

## Grundbegraber i sandsynlighedsregningen

Larsen, Jørgen

*Publication date:*  
1988

*Document Version*  
Også kaldet Forlagets PDF

*Citation for published version (APA):*  
Larsen, J. (1988). *Grundbegraber i sandsynlighedsregningen*. Roskilde Universitet. Tekster fra IMFUFA Nr. 166

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

### Take down policy

If you believe that this document breaches copyright please contact [rucforsk@ruc.dk](mailto:rucforsk@ruc.dk) providing details, and we will remove access to the work immediately and investigate your claim.

**TEKST NR 166**

**1988**

**Grundbegreber  
i  
Sandsynlighedsregningen**

Jørgen Larsen

IMFUFA  
Roskilde Universitetscenter

September 1988

**TEKSTER fra**

**IMFUFA**

**ROSKILDE UNIVERSITETSCENTER**

INSTITUT FOR STUDIET AF MATEMATIK OG FYSIK SAMT DERES  
FUNKTIONER I UNDERVISNING, FORSKNING OG ANVENDELSER

IMFUFA, Roskilde Universitetscenter, Postboks 260, DK-4000 Roskilde.

Jørgen Larsen: **Grundbegreber i Sandsynlighedsregningen**

IMFUFA-tekst 166/1988

iv + 89 sider

ISSN 0106-6242

---

*Grundbegreber i Sandsynlighedsregningen* er en præsentation af visse grundlæggende dele af sandsynlighedsregningen der har betydning i statistik; teksten kan bl.a. anvendes i forbindelse med *Basisstatistik*-noterne (IMFUFA-tekst 167 A&B).

Der er ikke tale om en lærebogsfremstilling der starter helt fra bunden. Hensigten er snarere at tilbyde den læser der tidligere har lært en smule sandsynlighedsregning en genopfriskning af nogle hovedpunkter i et måske anderledes sprog, suppleret med enkelte resultater som en sådan læser næppe har stiftet bekendtskab med.

Der er ikke lagt vægt på den aksiomatiske side af sandsynlighedsregningen, men på at bibringe læseren en såvel intuitiv som matematisk forståelse af de grundlæggende begreber; i den anledning er teksten forsynet med adskillige figurer.

# Forord

Dette hæfte præsenterer visse grundlæggende dele af sandsynlighedsregningen af betydning i statistik og kan bl.a. anvendes i forbindelse med *Basisstatistik*-noterne (IMFUFA-tekst 167 A&B).

Der er ikke tale om en lærebogsfremstilling der starter helt fra bunden. Hensigten er snarere at tilbyde den læser der tidligere har lært en smule sandsynlighedsregning en genopfriskning af nogle hovedpunkter i et måske anderledes sprog, suppleret med enkelte resultater som læseren næppe har stiftet bekendtskab med.

Læseren bør under alle omstændigheder sætte sig ind i Kapitel 1, afsnittene 2.1 - 2.4 i Kapitel 2, og Kapitel 3. Den matematisk interesserede læser bør desuden studere resten af Kapitel 2, samt Kapitel 4.

Kapitel 4 adskiller sig fra de øvrige afsnit dels ved at være væsentlig mere "matematisk", dels ved ikke at behandle generelle emner, men derimod, som et slags eksempel på brugen af de forrige afsnits teknikker og begreber, at bevise forskellige grundlæggende og vigtige resultater vedrørende normalfordelingen.

# Indhold

<b>Indledning</b>	<b>1</b>
<b>1 Sandsynligheder på endelige mængder</b>	<b>3</b>
1.1 De grundlæggende begreber . . . . .	3
1.2 Transformation og betingning . . . . .	7
1.3 Sammensatte forsøg . . . . .	14
<b>2 Sandsynligheder på <math>\mathbf{R}</math> og på <math>\mathbf{R}^n</math>.</b>	<b>21</b>
2.1 Diskrete fordelinger . . . . .	22
2.2 Kontinuerte fordelinger . . . . .	24
2.3 Fordelingsfunktionen . . . . .	28
2.4 Fraktiler . . . . .	31
2.5 Flerdimensionale fordelinger . . . . .	33
2.6 Transformation af kontinuerte fordelinger . . . . .	37
2.7 Transformation af flerdimensionale variable . . . . .	43
2.8 Betingede fordelinger . . . . .	50
<b>3 Middelværdi, varians, standardafvigelse</b>	<b>53</b>
3.1 Middelværdi . . . . .	53
3.2 Regneregler for middelværdier . . . . .	61
3.3 Varians . . . . .	62
<b>4 Nogle resultater om normalfordelingen</b>	<b>67</b>
4.1 Normalfordelingen . . . . .	67
4.2 $\chi^2$ -fordelingen . . . . .	74
4.3 Summer af normalfordelte variable . . . . .	79
4.4 Fordelingen af $\bar{X}$ og $s^2$ . . . . .	81
4.5 $t$ -fordelingen, $F$ -fordelingen . . . . .	83
<b>Liste over eksempler</b>	<b>87</b>

**Stikordsregister**

**88**

# Indledning

Sandsynlighedsregning er en matematik-disciplin der beskæftiger sig med en matematisk formalisering og kvantificering af tilfældighed. Sandsynlighedsregningen bygger på en *aksiomatisering* af det matematiske sandsynlighedsbegreb, og den beskæftiger sig med at *bevise* sætninger om egenskaber ved sandsynlighed og ved formelle sandsynlighedsmodeller. Noget af det interessante ved sandsynlighedsregningen ligger i, at ved hjælp af den kan man opbygge formelle modeller af tilfældighedsfænomener fra virkeligheden; de sætninger man kan bevise om de formelle modeller kan så oversættes til udsagn om de virkelige fænomeners opførsel.

Sandsynlighedsteoretikere bruger tit eksempler med terninger og mønter der kastes, ikke fordi det er det eneste man kan beskrive ved hjælp af sandsynlighedsregningen, men fordi mønt- og terningekast er sådan nogen gode håndgribelige tilfældighedsfænomener. Tag for eksempel terningekast: Hvis man kaster en almindelig terning mange gange, får man en sekser i ca. en sjettedel af kastene; det er en erfaringssag der intet har med matematik at gøre. Men om den formelle sandsynlighedsmodel for terningekast (som vi møder senere) kan man *bevise*, at "med sandsynlighed 1" vil den relative hyppighed af seksere gå mod  $1/6$  når antallet af kast vokser ud over alle grænser (ifølge *Store Tals Stærke Lov*). Endvidere kan man *bevise* resultater om, hvor meget man skal forvente at den faktisk observerede relative hyppighed af seksere kan afvige fra  $1/6$  når man foretager for eksempel 100 kast, og det viser sig at disse teoretisk forudsagte afvigelser stemmer overens med det man kan gå ud og observere. — At man sådan kan bevise noget der svarer til erfaringsmæssige kendsgerninger, skal tages som et tegn på, at sandsynlighedsregningen faktisk er velegnet til at beskrive virkelighedens tilfældighedsfænomener.

Som det vil fremgå af de øvrige dele af *Basisstatistik*-serien er sta-

tistiske modeller langt fra altid naturalistiske modeller der foregiver at efterligne faktiske tilfældighedsfænomener fra virkeligheden; lige så tit er det modelbyggeren der afgør, hvad der skal anses for tilfældigt. Statistikeren benytter sandsynlighedsmodeller til at beskrive sine tilfældigheder, hvadenten disse er konkrete eller de vedrører stikprøver fra "hypotetiske uendelige populationer". En pointe ved at formulere statistiske modeller i matematiksprø er, at man så kan bruge matematikken (sandsynlighedsregningen) til at bevise sætninger om egenskaber ved modellerne og ved de metoder man anvender for at analysere modellerne.

Selv om der altid kan ligge en intellektuel tilfredsstillelse i at udlede resultater om en formel model, så er meningen dog også at resultaterne skal oversættes til udsagn om det som modellen er en model for. Når vi skal oversætte sandsynlighedsmodellers udsagn til udsagn om virkeligheden, vil vi bruge *hyppighedsfortolkningen af sandsynligheder*:

Sandsynligheder skal fortolkes som relative hyppigheder i det lange løb<sup>1</sup>.

---

<sup>1</sup>Bemærk dog, at "skal fortolkes som" ikke er det samme som "er det samme som".



# Kapitel 1

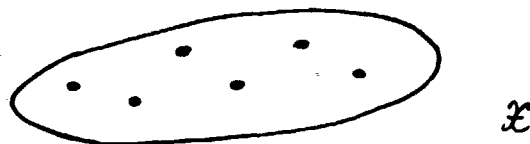
## Sandsynligheder på endelige mængder

Dette kapitel giver en elementær indføring til det matematiske begreb en sandsynlighedsfordeling og dertil knyttede begrebsdannelser, bl.a. stokastiske variable, og til transformationer og betingede fordelinger.

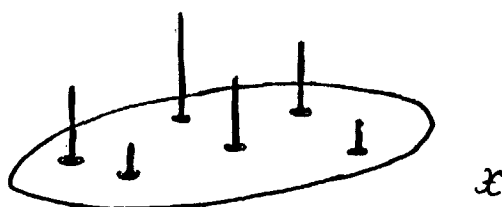
### 1.1 De grundlæggende begreber

Det første skridt i opbygningen af en sandsynlighedsmodel for et tilfældighedsfænomen er simpelthen at afgrænse de mulige resultater af fænomenet og formalisere dem til en mængde (i matematisk forstand) som traditionelt kaldes *udfaldsrummet* og hvis elementer tilsvarende kaldes *udfald*. Indtil videre vil vi udelukkende se på sandsynlighedsmodeller med et *endeligt* udfaldsrum, dvs. et udfaldsrum som er en endelig mængde. Et typisk udfaldsrum kan hedde  $\mathcal{X} = \{x_1, x_2, \dots, x_m\}$ , svarende til at de mulige udfald er  $x_1, x_2, \dots, x_m$ ; vi vil illustrere  $\mathcal{X}$  som i Figur 1.1.

Sandsynlighedsmodellen skal beskrive hvordan den samlede sandsynlighedsmasse 1 er fordelt blandt de enkelte udfald, og derfor består næste skridt i opbygningen af modellen i, at der til hvert udfald  $x$  knyttes et tal  $f(x) \geq 0$  kaldet *sandsynligheden for  $x$* , således at disse sandsynligheder summerer til 1, dvs.  $f(x_1) + f(x_2) + \dots + f(x_m) = 1$ .



Figur 1.1: Et endeligt udfaldsrum.



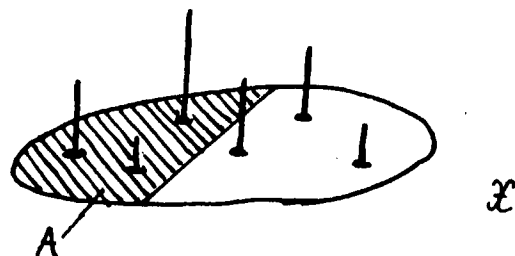
Figur 1.2: Et endeligt udfaldsrum med sandsynlighedspinde.

Sagt på en anden måde skal vi angive en funktion  $f$  som afbilder udfaldsrummet  $\mathcal{X}$  ind i  $[0, +\infty[$  og som opfylder betingelsen

$$\underbrace{f(x_1) + f(x_2) + \dots + f(x_m)}_{= \sum_{x \in \mathcal{X}} f(x)} = 1.$$

En sådan funktion  $f$  hedder en *sandsynlighedsfunktion*, og den beskriver en *sandsynlighedsfordeling* på  $\mathcal{X}$ . Figur 1.2 er en illustration af en sandsynlighedsfordeling; de enkelte pindes længde  $f(x)$  angiver de forskellige udfalds sandsynlighed, og de summerer til 1.

I den matematiske formalisme er man nu i stand til at tale om sandsynligheden for et udfald  $x$ , nemlig  $f(x)$ . Man vil imidlertid også gerne kunne tale om sandsynligheder for andet end enkeltudfald. Hvis  $A$  er en mængde af udfald, en såkaldt *hændelse*, dvs.  $A$  er en delmængde af udfaldsrummet, så skal "sandsynligheden for  $A$ " betyde summen af sandsynlighederne for enkeltudfaldene i  $A$ , se Figur 1.3. Traditionelt



Figur 1.3: Et udfaldsrum med en hændelse  $A$ .

forkorter man "sandsynligheden for  $A$ " til  $P(A)$ <sup>1</sup>. At sandsynligheden for  $A$ , dvs.  $P(A)$ , er lig summen af sandsynlighederne  $f(x)$  for de enkelte udfald  $x \in A$  skrives

$$P(A) = \sum_{x \in A} f(x).$$

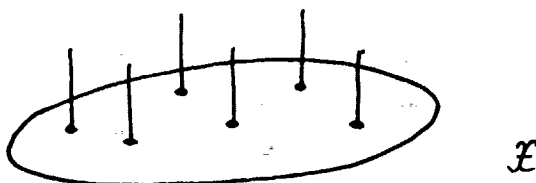
Man kan opfatte  $P$  som en funktion der til hver hændelse  $A$  knytter sandsynligheden  $P(A)$  for at  $A$  indtræffer. Denne funktion hedder en *sandsynlighedsfordeling* (på  $\mathcal{X}$ ).<sup>2</sup>

### Eksempel 1.1. Terningekast

For det tilfældighedsfænomen der består i at kaste en almindelig terning én gang og se hvor mange øjne den viser, er udfaldsrummet  $\mathcal{X} = \{1, 2, 3, 4, 5, 6\}$ . Hvis der ikke er nogen fiks-fakserier ved terningen, så skal alle udfaldene vel have samme sandsynlighed, og da sandsynlighederne skal summere til 1, må hvert enkelt udfald have sandsynlighed  $f(x) = 1/6$ . (Den derved beskrevne sandsynlighedsfordeling kaldes af nærliggende grunde for ligefordelingen på  $\mathcal{X} = \{1, 2, 3, 4, 5, 6\}$ .) (Se Figur 1.4.)

<sup>1</sup> $P$  kommer fra engelsk *probability* eller fransk *probabilité* eller latin *probabilitas*

<sup>2</sup>Læseren kan med rette undre sig over, at det er  $P$  og ikke  $f$  der har betegnelsen sandsynlighedsfordeling. Anledningen til at sprogbrugen er som den er skal søges i problemer (og deres løsning) ved indførelsen af sandsynligheder på uendelige abstrakte mængder, et emne vi ikke kommer ind på her.



Figur 1.4: Lige fordelingen. Hver pind har længden  $1/6$ .

Sandsynligheden for at få et ulige antal øjne er  $P(A)$ , hvor  $A = \{1, 3, 5\}$ . Man finder at

$$\begin{aligned}
 P(A) &= \sum_{x \in A} f(x) \\
 &= f(1) + f(3) + f(5) \\
 &= 1/6 + 1/6 + 1/6 \\
 &= 1/2.
 \end{aligned}$$

□

Der er tit behov for en bekvem og naturlig betegnelse for et udfald valgt tilfældigt med hensyn til en nærmere angivet sandsynlighedsfordeling. Til det formål bruger man såkaldte *stokastiske variable*<sup>3</sup>. Stokastiske variable betegnes ofte med store bogstaver ( $X, Y, Z, \dots$ ). Man bruger dem på den måde, at man starter med at oplyse, at f.eks.  $X$  er en stokastisk variabel med den og den fordeling; det betyder, at  $X$  fremover betegner *et udfald valgt tilfældigt med hensyn til denne fordeling*. Senere kan man så lade  $X$  indgå i alle udtryk hvor det er tilladt at indsætte punkter  $x$  fra  $\mathcal{X}$ . Eksempelvis kan man tale om sandsynligheden for at  $X \in A$  eller for at  $X = x_3$ . Hvis udfaldsrummet er en mængde af tal der kan adderes, så kan man tale om f.eks.  $X_1 + X_2$  hvis  $X_1$  og  $X_2$  er erklæret som stokastiske variable.

Man plejer at skrive  $P(\dots)$  som forkortelse for "sandsynligheden for at  $\dots$  indtræffer"; f.eks. betyder  $P(X \in A)$  sandsynligheden for at  $X$  tilhører  $A$ .

<sup>3</sup>Stokastisk betyder her noget i retning af "bestemt af tilfældigheder, som følger en vis fordeling".

## 1.2 Transformation og betingning

Hvis man har med et tilfældighedsfænomen at gøre (eller blot et eksperiment hvis resultat ikke kan forudsiges eksakt), kan man repræsentere udfaldet af det med en stokastisk variabel  $X$  med en bestemt sandsynlighedsfordeling. Nu kan det træffe sig så, at det ikke er  $X$  i sig selv der har den egentlige interesse, men derimod en vis funktion  $t(X)$  af  $X$ . (Eksempel: man er ikke interesseret i udfaldet af hvert enkelt af 100 møntkast, men kun i det samlede antal Krone.) Derved bliver der en opgave der hedder: bestem fordelingen af  $t(X)$ . Denne opgave løses generelt under overskriften *transformation af sandsynlighedsfordelinger*. — Det kan imidlertid også ske at man ikke kan få værdien af selve  $X$  at vide, men kun værdien af en vis funktion  $t(X)$  af  $X$  (man får at vide at modspilleren har "tre Hjerter"). Derved opstår en opgave af typen: når det er oplyst at  $t(X)$  har værdien  $y$ , hvad kan man så sige om  $X$ ? Denne opgave besvares formelt ved at bestemme en *betinget fordeling af  $X$  givet at  $t(X) = y$* .

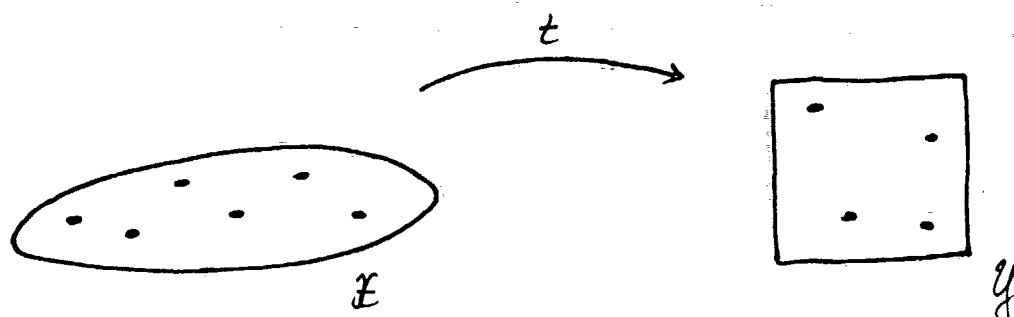
### Transformation

Først omtales *transformation af sandsynlighedsfordelinger*. I det matematiske set-up er situationen den, at der foreligger to mængder  $\mathcal{X}$  og  $\mathcal{Y}$  og en afbildning (transformation)  $t : \mathcal{X} \rightarrow \mathcal{Y}$ , jf. Figur 1.5. På  $\mathcal{X}$  er der en sandsynlighedsfordeling samt en stokastisk variabel  $X$  som følger denne fordeling. Opgaven er da at finde fordelingen af den stokastiske variabel  $Y = t(X)$ , dvs. at finde  $P(Y = y)$  for ethvert  $y \in \mathcal{Y}$ , jf. Figur 1.6.

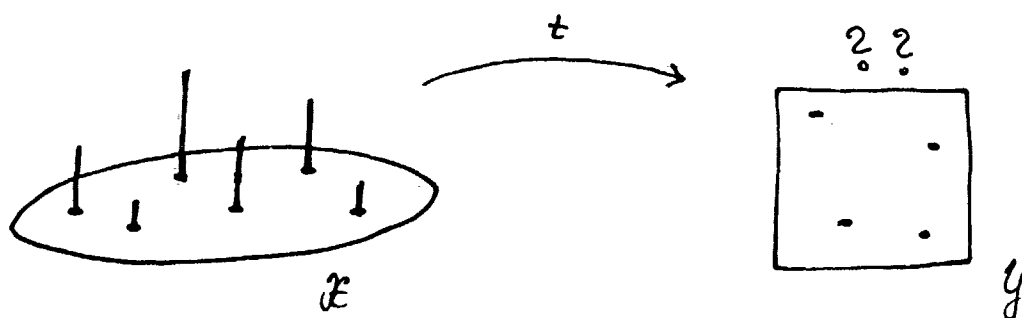
Løsningen er ligetil: Da  $Y$  er defineret til at være  $t(X)$ , er  $Y = y$  ensbetydende med at  $t(X) = y$ , dvs. med at  $X \in \{x : t(x) = y\}$ , så derfor er

$$\begin{aligned} P(Y = y) &= P(X \in \{x : t(x) = y\}) \\ &= \sum_{x : t(x)=y} f(x), \end{aligned} \tag{1.1}$$

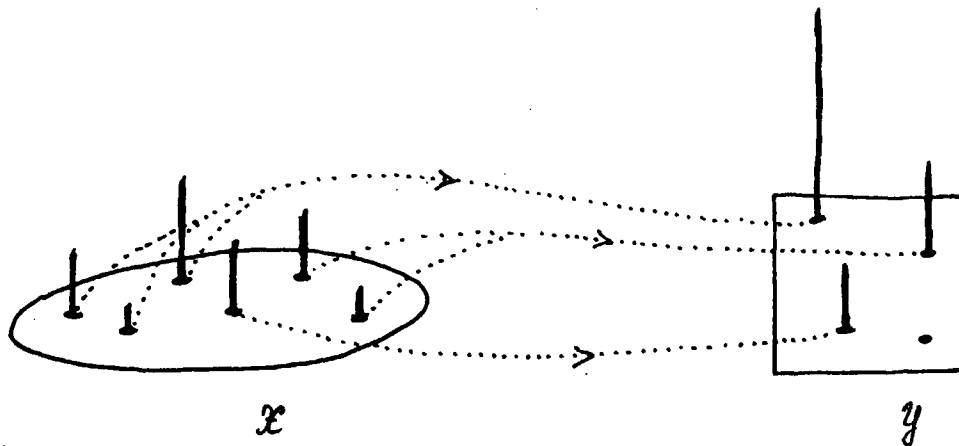
hvor  $f$  er sandsynlighedsfunktionen for  $X$  og hvor symbolet " $\sum_{x : t(x)=y}$ " betyder "summation over de  $x$  for hvilke  $t(x) = y$ ".



Figur 1.5: En transformation  $t$  fra  $X$  til  $Y$ .



Figur 1.6: Sandsynlighedsfordelingen på  $X$  skal transformeres.



Figur 1.7: Transformation af en sandsynlighedsfordeling.

Transformationen  $t$ , der i første omgang afbilder punkter i  $\mathcal{X}$  over i punkter i  $\mathcal{Y}$ , kan på denne måde også afbilde en sandsynlighedsfordeling på  $\mathcal{X}$  over i en sandsynlighedsfordeling på  $\mathcal{Y}$ , jf. Figur 1.7.

### Eksempel 1.2. Terningkast, fortsat

Lad som hidtil i terningekasteksemplet

$$\mathcal{X} = \{1, 2, 3, 4, 5, 6\}$$

og indfør desuden

$$\mathcal{Y} = \{\text{"delelig med 3"}, \text{"ikke delelig med 3"}\}.$$

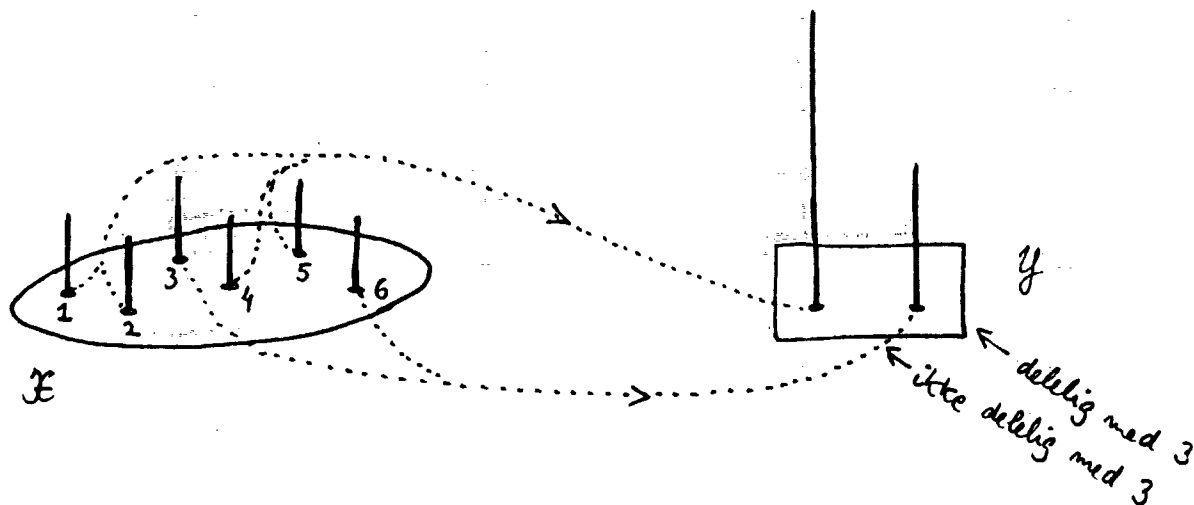
Indfør endelig afbildningen  $t : \mathcal{X} \rightarrow \mathcal{Y}$  givet ved at

$$t(1) = t(2) = t(4) = t(5) = \text{"ikke delelig med 3"}$$

og

$$t(3) = t(6) = \text{"delelig med 3"}.$$

Den transformerede fordeling er her bestemt ved at



Figur 1.8: Transformationen af sandsynlighedsfordelingen i Eksempel 1.2.

$$P(\text{"antal øjne deleligt med 3"}) = 2/6 = 1/3$$

og

$$P(\text{"antal øjne ikke deleligt med 3"}) = 4/6 = 2/3,$$

jf. Figur 1.8.

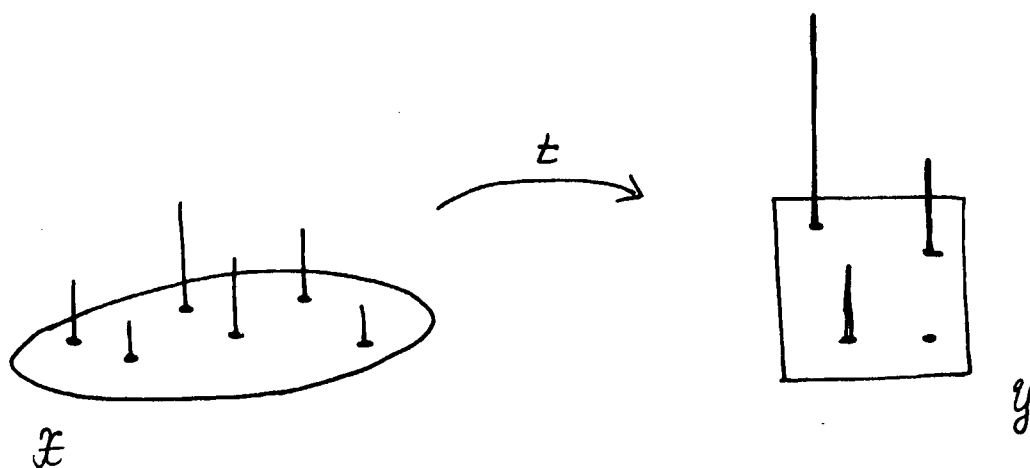
□

## Betingning

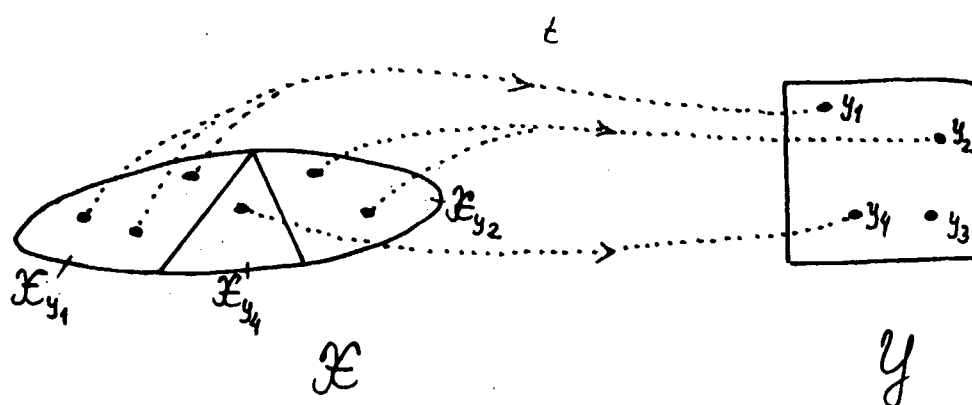
Vi går herefter over til spørgsmålet om at bestemme *betingede fordelinger*. Situationen er stadig den (jf. Figur 1.9), at der er to mængder  $\mathcal{X}$  og  $\mathcal{Y}$ , en afbildning  $t : \mathcal{X} \rightarrow \mathcal{Y}$  og en sandsynlighedsfordeling for den stokastiske variabel  $X$  på  $\mathcal{X}$ . Det er denne gang givet, at den stokastiske variabel  $Y = t(X)$  har værdien  $y$ . Opgaven er at finde ud af, hvad man på denne baggrund kan sige om  $X$ .

Man kan først konstatere at mulighedsområdet for  $X$  er blevet indskrænket:  $X$  må nødvendigvis antage en værdi som ved  $t$  afbildes over i  $y$ , dvs.  $X$  må tilhøre delmængden  $\mathcal{X}_y = \{x : t(x) = y\}$  af  $\mathcal{X}$ , jf. Figur 1.10. Endvidere må det være sådan, at hvis f.eks.  $x_1$  og  $x_2$  begge





Figur 1.9: Transformeret sandsynlighedsfordeling.



Figur 1.10:  $t$  bestemmer en klassesdeling af  $X$ .

afbilledes over i det samme  $y$ , dvs.  $x_1$  og  $x_2$  begge tilhører  $\mathcal{X}_y$ , så kan oplysningen  $t(X) = y$  ikke ændre på *forholdet* mellem sandsynlighederne for at  $X = x_1$  hhv.  $X = x_2$ . Det vil altså sige, at forholdet mellem de betingede sandsynligheder givet at  $t(X) = y$  for at  $X = x_1$  hhv.  $X = x_2$  er lig med forholdet mellem de oprindelige ubetingede sandsynligheder for at for at  $X = x_1$  hhv.  $X = x_2$ . Det skrives lidt mere overskueligt som

$$\frac{P(X = x_1 | t(X) = y)}{P(X = x_2 | t(X) = y)} = \frac{P(X = x_1)}{P(X = x_2)},$$

hvor  $P(X = x | t(X) = y)$  er standardbetegnelsen for "den betingede sandsynlighed for at  $X = x$  givet at  $t(X) = y$ ".

Der må derfor gælde at

$$P(X = x | t(X) = y) = \begin{cases} k \times P(X = x) & \text{hvis } (x) = y, \\ 0 & \text{ellers,} \end{cases}$$

hvor  $k$  er en konstant der bestemmes således at de betingede sandsynligheder givet at  $t(X) = y$  summerer til 1. Nu er

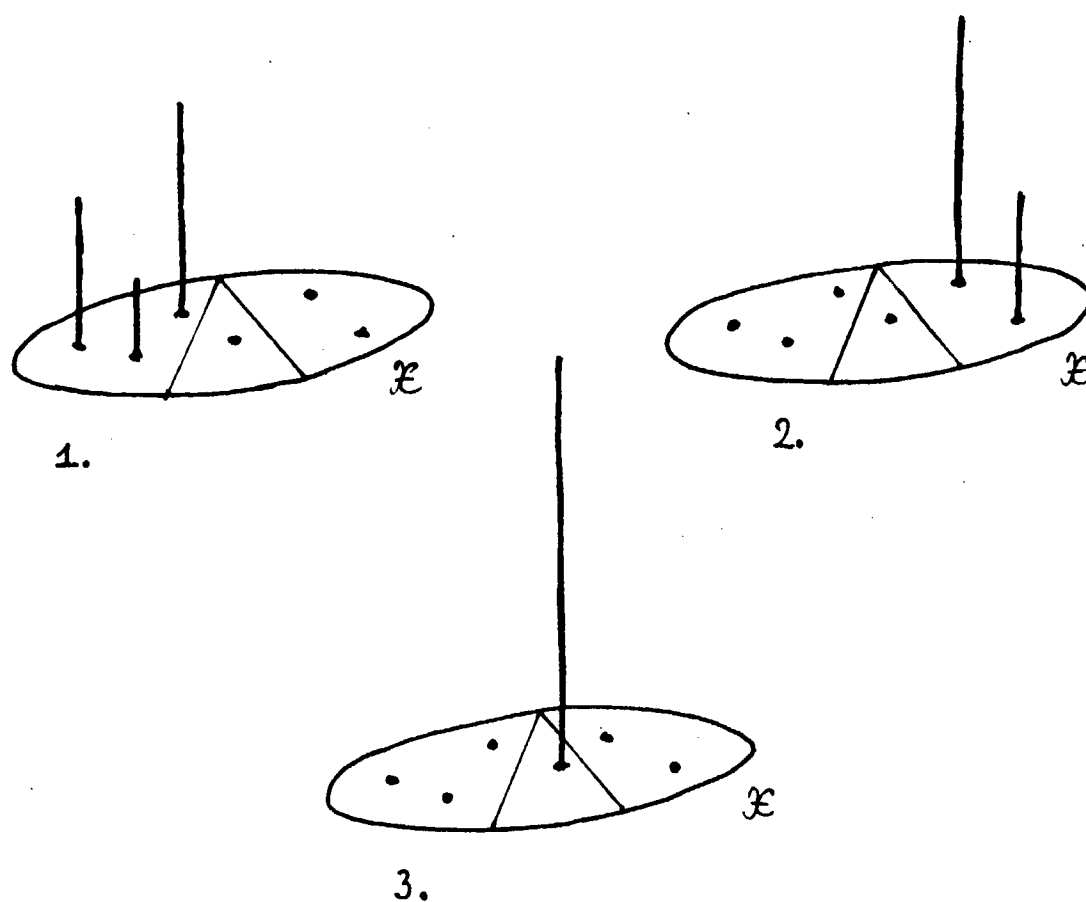
$$\begin{aligned} \sum_{x \in \mathcal{X}} P(X = x | t(X) = y) &= \sum_{x \in \mathcal{X}_y} k \times P(X = x) \\ &= k \times \sum_{x \in \mathcal{X}_y} P(X = x) \\ &= k \times P(X \in \mathcal{X}_y) \\ &= k \times P(t(X) = y), \end{aligned}$$

så konstanten skal være  $k = 1/P(t(X) = y)$ . Vi har hermed bestemt *den betingede sandsynlighed for at  $X = x$  givet at  $t(X) = y$* :

$$P(X = x | t(X) = y) = \begin{cases} \frac{P(X = x)}{P(t(X) = y)} & \text{hvis } t(X) = y \\ 0 & \text{ellers.} \end{cases}$$

Den sandsynlighedsfordeling på  $\mathcal{X}$  som har sandsynlighedsfunktionen  $f: x \mapsto P(X = x | t(X) = y)$  hedder *den betingede fordeling af  $X$  givet at  $t(X) = y$* <sup>4</sup>. Figur 1.11 viser et eksempel på betingede fordelinger.

<sup>4</sup>Bemærkning: I udledningerne (og i den resulterende formel) er det stiltiende forudsat at  $P(t(X) = y)$  ikke er 0; hvis  $P(t(X) = y)$  er 0, så kan man som betinget fordeling af  $X$  givet at  $t(X) = y$  vælge en hvilken som helst sandsynlighedsfordeling på  $\mathcal{X}_y$ .

**Figur 1.11:**

- 1) Den betingede fordeling af  $X$  givet at  $t(X) = y_1$ .
- 2) Den betingede fordeling af  $X$  givet at  $t(X) = y_2$ .
- 3) Den betingede fordeling af  $X$  givet at  $t(X) = y_4$ .

**Eksempel 1.3. Terningekast, fortsat**

Sandsynligheden for at få en toer givet at man får et antal øjne som ikke er deleligt med 3 er

$$\frac{P(2 \text{ øjne})}{P(\text{ikke delelig med } 3)} = \frac{1/6}{2/3} = \frac{1}{4}.$$

□

Vi har set at man ud fra  $X$  og  $t$  kan bestemme dels fordelingen af  $t(X)$ , dels de betingede fordelinger af  $X$  givet at  $t(X) = y$ . Imidlertid kan man også regne tilbage: hvis man kender fordelingen af  $t(X)$  og den betingede fordeling af  $X$  givet at  $t(X) = y$  for ethvert  $y \in t(\mathcal{X})$ , så kan man ud fra disse finde den oprindelige fordeling af  $X$ . Det ses ved at gange med  $P(t(X) = y)$  på begge sider af lighedstegnet i (1.2) og skrive  $t(x)$  i stedet for  $y$ ; så får vi udtrykt  $P(X = x)$  ved hjælp af de betingede sandsynligheder og de transformerede sandsynligheder:

$$P(X = x) = \quad (1.2)$$

$$P(X = x \mid t(X) = t(x)) \times P(t(X) = t(x)).$$

**1.3 Sammensatte forsøg**

De tilfældighedsfænomener man har med at gøre er meget ofte fænomener der med rimelighed kan opfattes som sammensat af et antal simple del-fænomener (man kaster mønten eller terningen mange gange; man foretager mange målinger på hvert forsøgsdyr; man måler på mange forsøgsdyr etc.). Fra et modelbygningsmæssigt synspunkt er det derfor bekvemt med nogle teknikker til opbygning af større modeller ud fra simple delmodeller.

Lad os sige at udfald af det samlede forsøg eller fænomen indeholder  $n$  komponenter  $x_1, x_2, \dots, x_n$ , og at udfaldsrummet tilsvarende kan skrives som en produktmængde

$$\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \times \dots \times \mathcal{X}_n,$$

dvs. at et udfald kan skrives som  $\mathbf{x} = (x_1, x_2, \dots, x_n)$ . En stokastisk variabel til at repræsentere det tilfældige udfald af forsøget kan tilsvarende skrives som  $\mathbf{X} = (X_1, X_2, \dots, X_n)$ ; en sådan stokastisk variabel vil man kalde en *n-dimensional stokastisk variabel*.

Den stokastiske variabel  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  har en sandsynlighedsfunktion  $f(x_1, x_2, \dots, x_n)$  der beskriver den *simultane fordeling* af  $X_1, X_2, \dots, X_n$ . Den dertil hørende *marginale fordeling* af f.eks.  $X_2$  er fordelingen af  $x_2$  alene uden hensyn til de øvrige komponenter. Fra et matematisk synspunkt er den marginale fordeling af f.eks.  $X_2$  lig med den transformerede fordeling af  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  ved transformationen  $t: (x_1, x_2, \dots, x_n) \mapsto x_2$  og findes derfor ved brug af (1.1) der viser at man får sandsynlighedsfunktionen  $f_2(x_2)$  for marginalfordelingen af  $X_2$  ved at "summere  $x_1, x_3, x_4, \dots, x_n$  ud":

$$\begin{aligned} f_2(x_2) &= P(t(\mathbf{X}) = x_2) \\ &= \sum_{\mathbf{x}: t(\mathbf{x})=x_2} f(x_1, x_2, \dots, x_n) \\ &= \sum_{x_1} \sum_{x_3} \sum_{x_4} \dots \sum_{x_n} f(x_1, x_2, \dots, x_n). \end{aligned}$$

#### Eksempel 1.4. Marginal og simultan fordeling

Antag at  $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 = \{1, 2, 3\} \times \{1, 2\}$ , hvilket illustreres således:

2	•	•	•
1	•	•	•
	1	2	3

Der er på  $\mathcal{X}$  en sandsynlighedsfordeling med sandsynlighederne

0.2	0.4	0.1
0.1	0.1	0.1

Det vil f.eks. sige at  $f(1, 2) = 0.2$ , hvor  $f$  er sandsynlighedsfunktionen for  $\mathbf{X}$ .

Her har vi den simultane og de marginale fordelinger:

den simultane  
fordeling af  $X_1 \rightarrow$ 

0.2	0.4	0.1
0.1	0.1	0.1

0.7
0.3

 $\leftarrow$  den marginale  
og  $X_2$  fordeling af  $X_2$

den marginale  
fordeling af  $X_1 \rightarrow$ 

0.3	0.5	0.2
-----	-----	-----

1
---

 $\leftarrow$  totalsummen

Derved er begreberne *marginal og simultan fordeling* illustreret. Af hensyn til det følgende bestemmes også de betingede fordelinger af  $X_1$  givet  $X_2$ <sup>5</sup>. Fordelingerne er

$$\begin{bmatrix} 2/7 & 4/7 & 1/7 \end{bmatrix} \quad \text{når } X_2 = 2,$$

$$\begin{bmatrix} 1/3 & 1/3 & 1/3 \end{bmatrix} \quad \text{når } X_2 = 1.$$

De to betingede fordelinger er forskellige, og det betyder at  $X_2$  indeholder information om  $X_1$ : Givet at  $X_2 = 1$  er alle  $x_1$ -værdier lige sandsynlige, men givet at  $X_2 = 2$  er f.eks. udfaldet  $x_1 = 2$  fire gange så sandsynlig som udfaldet  $x_1 = 3$ . — Man taler om at  $X_1$  og  $X_2$  er stokastisk afhængige.  $\square$

Når man skal opbygge modeller for sammensatte fænomener ud fra modeller for delfænomener, er det lettest at gøre i de situationer hvor det er sådan, at udfaldet af ét bestemt delfænomen ikke påvirkes af hvad udfaldene af de andre delfænomener nu tilfældigvis blev. Man taler om at delfænomenerne er *stokastisk uafhængige*. Den formelle definition af stokastisk uafhængighed er:

**Definition:** Stokastiske variable  $X_1, X_2, \dots, X_n$  er *stokastisk uafhængige*, når det for enhver af de variable gælder, at dens betingede fordelinger givet en eller flere af de øvrige variable kan vælges som dens ubetingede (marginale) fordeling.

Definitionen viser (forhåbentlig) at begrebet stokastisk uafhængighed er et fornuftigt begreb, men det fremgår ikke hvordan det kan være til gavn i modelbygningsmæssige sammenhænge. Det viser sig imidlertid, at når stokastiske variable er uafhængige, så kan deres simultane sandsynlighedsfunktion skrives som et produkt af de marginale sandsynlighedsfunktioner for de enkelte komponenter, og derfor er det let at opskrive den simultane sandsynlighedsfunktion når man kender de enkelte komponenters marginale sandsynlighedsfunktioner. Denne egenskab ved stokastisk uafhængighed er præciseret som

**Sætning 1.1** *Stokastiske variable  $X_1, X_2, \dots, X_n$  er stokastisk uafhængige, hvis og kun hvis deres simultane sandsynlighedsfunktion*

<sup>5</sup>Eller rettere: de marginale fordelinger af  $X_1$  i de betingede fordelinger af  $(X_1, X_2)$  givet  $X_2$ .

$f(x_1, x_2, \dots, x_n)$  kan skrives som et produkt af de enkelte komponenters marginale sandsynlighedsfunktioner, dvs. hvis og kun hvis

$$f(x_1, x_2, \dots, x_n) = f_1(x_1) \times f_2(x_2) \times \dots \times f_n(x_n),$$

hvor  $f_i(x_i)$  er den marginale sandsynlighedsfunktion for  $X_i$ .

Her er ideen til beviset for sætningen; for overskuelighedens skyld antager vi at  $n = 3$ :

1. Vi viser først "kun hvis", så det antages at  $X$ -erne er uafhængige. Ved hjælp af 1.2 fås omskrivningen

$$\begin{aligned} f(x_1, x_2, x_3) &= P(X_1 = x_1, X_2 = x_2, X_3 = x_3) \\ &= P(X_1 = x_1, X_2 = x_2, X_3 = x_3 \mid X_2 = x_2, X_3 = x_3) \\ &\quad \times P(X_2 = x_2, X_3 = x_3) \\ &= P(X_1 = x_1 \mid X_2 = x_2, X_3 = x_3) \\ &\quad \times P(X_2 = x_2, X_3 = x_3). \end{aligned}$$

Ifølge antagelsen om uafhængighed er den betingede sandsynlighed  $P(X_1 = x_1 \mid X_2 = x_2, X_3 = x_3)$  lig med den marginale sandsynlighed  $P(X_1 = x_1)$  som pr. definition er lig  $f_1(x_1)$ . Dermed har vi at

$$f(x_1, x_2, x_3) = f_1(x_1) \times P(X_2 = x_2, X_3 = x_3).$$

I denne formel omskrives sidste faktor ved at bruge samme trick en gang til:

$$\begin{aligned} P(X_2 = x_2, X_3 = x_3) &= P(X_2 = x_2 \mid X_3 = x_3) \times P(X_3 = x_3) \\ &= f_2(x_2) \times f_3(x_3), \end{aligned}$$

så alt i alt er

$$f(x_1, x_2, x_3) = f_1(x_1) \times f_2(x_2) \times f_3(x_3).$$

Det var ideen til beviset for den ene halvdel ("kun hvis") af sætningen. Den anden del ("hvis") forløber således:

2. Antag i denne omgang at den simultane sandsynlighedsfunktion er et produkt af de marginale sandsynlighedsfunktioner

$$f(x_1, x_2, x_3) = f_1(x_1) \times f_2(x_2) \times f_3(x_3) .$$

Da er f.eks. (jf. (1.2))

$$\begin{aligned} P(X_1 = x_1 \mid X_2 = x_2, X_3 = x_3) \\ &= P(X_1 = x_1, X_2 = x_2, X_3 = x_3 \mid X_2 = x_2, X_3 = x_3) \\ &= \frac{P(X_1 = x_1, X_2 = x_2, X_3 = x_3)}{P(X_2 = x_2, X_3 = x_3)} \\ &= \frac{f(x_1, x_2, x_3)}{\sum_{z_1} f(z_1, x_2, x_3)} \\ &= \frac{f_1(x_1) \times f_2(x_2) \times f_3(x_3)}{\sum_{z_1} f_1(z_1) \times f_2(x_2) \times f_3(x_3)} \\ &= \frac{f_1(x_1)}{\sum_{z_1} f_1(z_1)} \\ &= f_1(x_1)/1 \\ &= f_1(x_1) . \end{aligned}$$

Dermed er vist, at den betingede fordeling af  $X_1$  givet at  $X_2 = x_2$  og  $X_3 = x_3$  er lig med den marginale fordeling af  $X_1$ .

### Eksempel 1.5. *Opbygning af større modeller ud fra mindre*

Vi illustrerer hvordan man kan bruge Sætning 1.1 ved opbygning af større modeller ud fra mindre modeller med følgende (afgjort ikke særlig relevante) eksempel.

Som led i et spil kaster hver spiller en almindelig terning én gang og en lidt skæv mønt to gange. Udfaldsrummet for dette eksperiment er

$$\mathcal{X} = \{1, 2, 3, 4, 5, 6\} \times \{\text{Plat}, \text{Krone}\} \times \{\text{Plat}, \text{Krone}\} .$$

Sandsynlighedsfunktionen for terningekast-delen er

$$f_1(x_1) = 1/6 , \quad x_1 = 1, 2, 3, 4, 5, 6 .$$



Mønten er som nævnt lidt skæv, så den viser ikke Krone med sandsynlighed  $1/2$  men med en sandsynlighed som er lidt mindre, vistnok omkring  $2/5$ . Da talværdien ikke er helt kendt, nøjes vi dog med at kalde sandsynligheden for  $p$ . Sandsynlighedsfunktionen for det første møntkast er så

$$f_2(x_2) = \begin{cases} p & \text{når } x_2 = \text{Krone,} \\ 1 - p & \text{når } x_2 = \text{Plat,} \end{cases}$$

og sandsynlighedsfunktionen for det andet møntkast er

$$f_3(x_3) = \begin{cases} p & \text{når } x_3 = \text{Krone,} \\ 1 - p & \text{når } x_3 = \text{Plat.} \end{cases}$$

Det må være rimeligt at gå ud fra, at terningekastet og de to møntkast ikke påvirker hinanden, at de altså er stokastisk uafhængige. Så er den simultane sandsynlighedsfunktion produktet af de marginale sandsynlighedsfunktioner:

$$f(x_1, x_2, x_3) = f_1(x_1) \times f_2(x_2) \times f_3(x_3).$$

Eksempelvis er da sandsynligheden for, først at få en femmer, så Plat og til sidst Krone

$$\begin{aligned} f(5, \text{Plat}, \text{Krone}) &= 1/6 \times (1 - p) \times p \\ &= p(1 - p)/6. \end{aligned}$$

□

### Eksempel 1.6. Om fortolkningen af usædvanlige hændelser

Ved 100 kast med en terning har man ingen af gangene fået en sekser. Hvad skal man så mene om chancen for at kast nr. 101 går hen og giver en sekser?

Mange "almindelige mennesker" (og mange spillere!) mener, at chancen for en sekser i den situation er væsentlig større end  $1/6$  (det er noget med at Naturen har sparet en hel masse seksere sammen ...).

Sandsynlighedsteoretikeren mener at de enkelte kast er stokastisk uafhængige; i så fald er det komplet ligegyldigt hvad de første 100 kast gav, kast nr. 101 giver en sekser med sandsynlighed  $1/6$ .

Statistikeren mener også at kastene er stokastisk uafhængige. Det benytter han til at udregne, at for en almindelig terning er sandsynligheden for 0 seksere i 100 kast lig  $(5/6)^{100} = 1.2 \times 10^{-8}$ . I henhold til hyppighedsfortolkningen af sandsynligheder (side 2) gælder, at hvis vi gentager forsøget "100 kast med terningen" et stort antal gange, så vil hændelsen "0 seksere" indtræffe i ca. en ud af 100 millioner gange, i det lange løb. Statistikeren vil sige, at da hændelser med så lille sandsynlighed "ikke indtræffer i praksis", tyder observationerne på at der er noget suspekt ved terningen. Statistikeren vil derfor være tilbøjelig til at mene, at chancen for at kast nr. 101 giver en sekser er væsentlig mindre end  $1/6$ .  $\square$

## Kapitel 2

# Sandsynligheder på $\mathbb{R}$ og på $\mathbb{R}^n$ .

De sandsynlighedsmodeller som statistikeren har brug for er ingenlunde altid modeller hvor udfaldsrummet er en eller anden ustruktureret endelig mængde svarende til *kvalitative* observationer. Meget ofte ønsker man at formulere modeller for *kvantitative* observationer, typisk enten antal ("tællelital") eller målinger på en kontinuert måleskala.

Hvis man vil opstille en model for et fænomen hvor udfaldet er en *antalsvariabel*, så er der i almindelighed to muligheder for hvad udfaldsrummet  $\mathcal{X}$  bør være:

- $\mathcal{X}$  kan være en endelig mængde  $\{0, 1, 2, \dots, n\}$  (for en passende værdi af  $n$ ), svarende til at der er en kendt øvre grænse for de mulige observationer.
- $\mathcal{X}$  kan være den uendelige mængde

$$\mathbb{N}_0 = \{0, 1, 2, \dots\}$$

bestående af alle ikke-negative hele tal, svarende til at der principielt ikke er nogen øvre grænse for, hvilket udfald man kan observere.

I begge tilfælde ligger udfaldene adskilt fra hinanden, og man taler om et *diskret udfaldsrum* og om en *diskret stokastisk variabel*.

Når man derimod skal modellere et fænomen hvor udfaldet er en *kontinuert varierende størrelse*, en "måling" (f.eks. af en tid, højde,

længde, masse, koncentration etc.), så kan man ofte benytte en af følgende typer udfaldsrum:

- $\mathcal{X}$  kan være den reelle talakse

$$\mathbf{R} = ] - \infty, +\infty[ ,$$

svarende til at der ikke er nogen hverken øvre eller nedre grænse for de mulige observationer.

- $\mathcal{X}$  kan være et nedadtil begrænset delinterval af  $\mathbf{R}$ , typisk de positive reelle tal  $]0, +\infty[$ , svarende til at der er en kendt nedre grænse for de mulige observationer.
- $\mathcal{X}$  kan være et begrænset interval af  $\mathbf{R}$ , f.eks.  $[0, 1]$  eller  $[0, 2\pi[$ , svarende til at der er både en kendt øvre og en kendt nedre grænse for, hvilke observationer det er muligt at få.

I sådanne udfaldsrum gælder, at et givet udfald  $x$  har naboer vilkårligt tæt ved sig; udfaldsrummet er sammenhængende, *kontinuert*.

Vi skal nu se, hvordan man beskriver sandsynlighedsfordelinger for en stokastisk variabel  $X$  på de to forskellige slags udfaldsrum.

## 2.1 Diskrete fordelinger

Et udfaldsrum af typen  $\mathcal{X} = \{0, 1, 2, \dots, n\}$  er en endelig mængde, så sandsynligheder på et sådant udfaldsrum kan beskrives fuldstændig som i Kapitel 1, nemlig ved hjælp af en *sandsynlighedsfunktion*  $f$ :

$$f(x) = P(X = x), \quad x \in \mathcal{X} ,$$

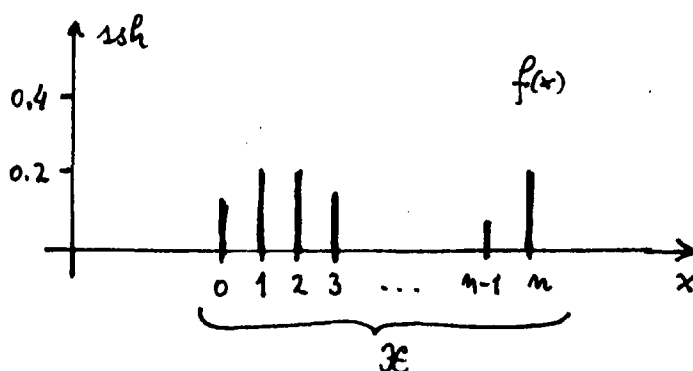
hvor  $f$  er ikke-negativ og summerer til 1:

$$f(x) \geq 0, \quad x \in \mathcal{X},$$

og

$$\sum_{x \in \mathcal{X}} f(x) = 1 .$$

Figur 2.1 viser et eksempel på en sandsynlighedsfunktion.



Figur 2.1: En sandsynlighedsfunktion  $f$  på  $\{0, 1, \dots, n\}$ .

Også hvis udfaldsrummet er  $\mathcal{X} = \mathbb{N}_0 = \{0, 1, 2, \dots\}$  beskriver man sandsynlighedsfordelinger ved en sandsynlighedsfunktion

$$f(x) = P(X = x), \quad x \in \mathcal{X}.$$

Her skal stadig gælde at

$$f(x) \geq 0, \quad x \in \mathcal{X},$$

og

$$\sum_{x \in \mathcal{X}} f(x) = 1,$$

men den sidste betingelse (at  $f$  skal summere til 1) er knap så uskyldig som før, for nu handler den om at summere uendelig mange tal.

### Eksempel 2.1. Den geometriske fordeling

Man slår Plat eller Krone med en mønt som er skæv og som giver Krone med sandsynlighed  $p$ . Lad  $X$  betegne det antal gange man slår Plat inden den første Krone.  $X$  er da en stokastisk variabel der i princippet kan antage en hvilken som helst ikke-negativ heltallig værdi, dvs. udfaldsrummet er  $\mathcal{X} = \mathbb{N}_0$ .

Man kan vise, at under de sædvanlige antagelser om møntkast er

$$\begin{aligned} P(X = x) &= P(\text{først } x \text{ gange Plat og så 1 gang Krone}) \\ &= p(1 - p)^x, \end{aligned}$$

dvs. fordelingen af den stokastiske variabel  $X$  er givet ved sandsynlighedsfunktionen

$$f(x) = p(1-p)^x, \quad x = 0, 1, 2, \dots \quad (2.1)$$

Denne fordeling hedder den geometriske fordeling med parameter  $p$ . Figur 2.2 vise et eksempel på en sandsynlighedsfunktion for en geometrisk fordeling.

Sandsynlighederne (2.1) summerer til 1,

$$\sum_{x=0}^{\infty} f(x) = 1,$$

forstået på den måde at

$$\lim_{N \rightarrow \infty} \sum_{x=0}^N f(x) = 1.$$

I det aktuelle eksempel er

$$\begin{aligned} \sum_{x=0}^N f(x) &= \sum_{x=0}^N p(1-p)^x \\ &= p \frac{1 - (1-p)^{N+1}}{1 - (1-p)} \\ &= 1 - (1-p)^{N+1} \end{aligned}$$

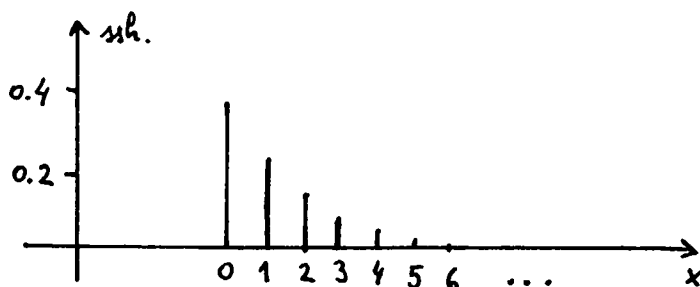
ifølge reglen om summation af en kvotientrække. Det ses at

$$\begin{aligned} \lim_{N \rightarrow \infty} \sum_{x=0}^N f(x) &= \lim_{N \rightarrow \infty} (1 - (1-p)^{N+1}) \\ &= 1. \end{aligned}$$

□

## 2.2 Kontinuerte fordelinger

Hvis udfaldsrummet er den reelle akse eller et delinterval af den reelle akse, så beskriver man en såkaldt *kontinuert* sandsynlighedsfordeling for



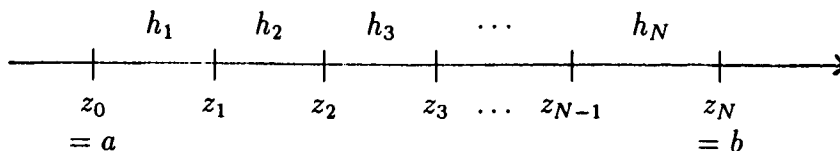
Figur 2.2: Sandsynlighedsfunktionen for den geometriske fordeling med  $p = 0.382$ .

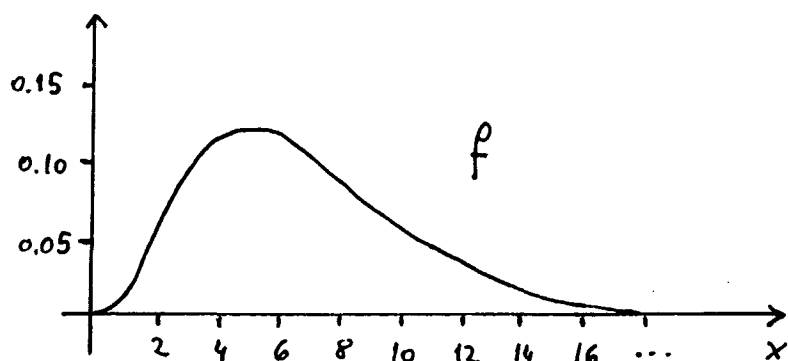
$X$  ved hjælp af en funktion der hedder en *sandsynlighedstæthedsfunktion*; denne fortæller, som navnet antyder, for hvert punkt  $x$  i udfaldsrummet hvor "tæt" sandsynlighedsmassen ligger netop dér. Formelt defineres sandsynlighedstæthedsfunktionen  $f(x)$  for  $X$  ved

$$f(x) = \lim_{h \searrow 0} \frac{P(X \text{ ligger i et } h\text{-interval omkring } x)}{h}.$$

Her er "et  $h$ -interval omkring  $x$ " et interval af længde  $h > 0$  omkring  $x$ . Definitionen betyder altså, at  $P(X \text{ ligger i et } h\text{-interval omkring } x)$  ca. er lig med  $f(x) \times h$  når  $h$  er nær 0. Figur 2.3 viser et eksempel på hvordan en sandsynlighedstæthedsfunktion kan se ud; bemærk i øvrigt, at sandsynlighedstæthedsfunktioner pr. definition altid er  $\geq 0$ .

Ud fra tæthedsfunktionen  $f$  kan man bestemme alle håndte sandsynligheder vedrørende  $X$ . Lad os prøve at bestemme  $P(a < X \leq b)$  hvor  $a \leq b$ : Del intervallet  $]a, b]$  op i  $N$  småintervaller af længder  $h_1, h_2, \dots, h_N$ :





Figur 2.3: En sandsynlighedstæthedsfunktion  $f$ .

Nu er den søgte sandsynlighed  $P(a < X \leq b)$  lig med summen af sandsynlighederne for at  $X$  ligger i de forskellige småintervaller, altså

$$P(X \in ]z_0, z_1]) + P(X \in ]z_1, z_2]) + \dots + P(X \in ]z_{N-1}, z_N]) .$$

Vælg et punkt  $x$  i hvert interval. Det punkt der vælges i det  $i$ -te interval betegnes  $x_i$ . Da interval nr.  $i$  er et  $h_i$ -interval omkring  $x_i$  er

$$P(X \in ]z_{i-1}, z_i]) \approx f(x_i) \times h_i .$$

Alt i alt er da

$$P(a < X \leq b) \approx f(x_1)h_1 + f(x_2)h_2 + \dots + f(x_N)h_N . \quad (2.2)$$

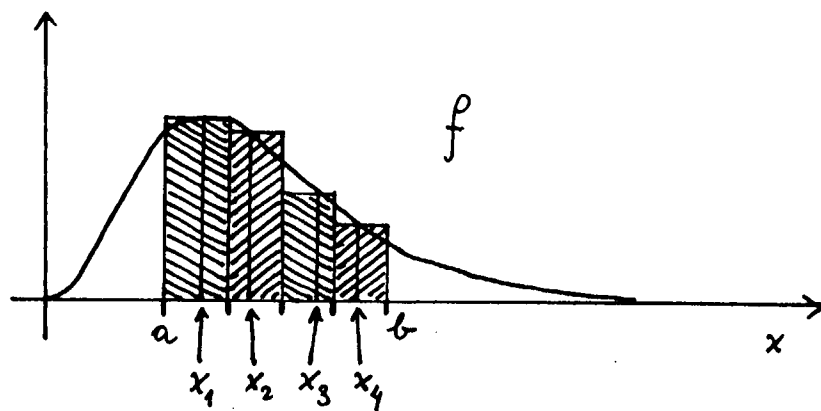
Hvad højresiden i denne formel er, illustreres i Figur 2.4 i tilfældet  $N = 4$ . Når man gør antallet  $N$  af delintervaller større og større og samtidig gør delintervallernes længder mindre og mindre, vil man i grænsen få selve arealet under kurven  $f$ , og vi har dermed fundet at

$$P(a < X \leq b) = \int_a^b f(x) dx ,$$

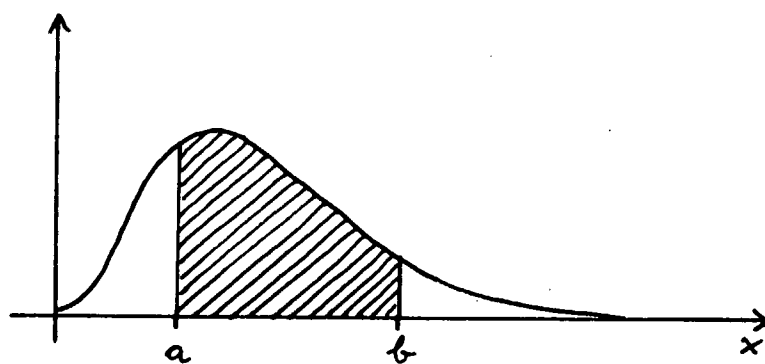
jf. Figur 2.5. Det forhold at  $X$  altid vil være beliggende et eller andet sted på talaksen indebærer at  $P(-\infty < X < +\infty)$  må være lig 1, og det betyder igen at  $f$  integrerer til 1:

$$\int_{-\infty}^{+\infty} f(x) dx = 1 .$$





Figur 2.4: Arealet af det skraverede område er højresiden i (2.2).



Figur 2.5: Arealet af det skraverede område er  $P(a < X \leq b) = \int_a^b f(x) dx$ .

Disse løse betragtninger kan (under forudsætning af passende antagelser om  $f$ ) bevises formelt, hvilket vi dog ikke vil gøre her. Vi nøjes med en oversigt over nogle vigtige resultater om henholdsvis sandsynlighedsfunktioner og sandsynlighedstæthedsfunktioner.

**I det diskrete tilfælde** gælder for en stokastisk variabel  $X$  med sandsynlighedsfunktion  $f$

1.  $f(x) = P(X = x)$ .
2.  $P(a < X \leq b) = \sum_{x=a+1}^b f(x)$ , når  $a < b$  er heltal.
3.  $\sum_x f(x) = 1$ .

**I det kontinuerte tilfælde** gælder for en stokastisk variabel  $X$  med sandsynlighedstæthedsfunktion  $f$

1.  $f(x) = \lim_{h \searrow 0} \frac{P(X \text{ ligger i et } h\text{-interval omkring } x)}{h}$ ,
2.  $P(a < X \leq b) = \int_a^b f(x) dx$ , når  $a \leq b$ .
3.  $\int_{-\infty}^{+\infty} f(x) dx = 1$ .

## 2.3 Fordelingsfunktionen

Alle slags sandsynlighedsfordelinger på den reelle akse eller delmængder af den reelle akse kan beskrives ved hjælp af deres *kumulerede fordelingsfunktion* (eller blot: *fordelingsfunktion*). Fordelingsfunktionen  $F$  for fordelingen af den stokastiske variabel  $X$  defineres ved

$$F(x) = P(X \leq x), \quad x \in \mathbf{R}. \quad (2.3)$$

Hvis  $X$  er heltallig, kan man udtrykke fordelingsfunktionen  $F$  ved sandsynlighedsfunktionen  $f$ , og omvendt:

$$\begin{aligned} F(x) &= \sum_{z \leq x} f(z), \\ f(x) &= F(x) - F(x-1). \end{aligned} \quad (2.4)$$

Hvis  $X$  er kontinuert ser sammenhængen mellem fordelingsfunktionen  $F$  og sandsynlighedstæthedsfunktionen  $f$  sådan ud<sup>1</sup>:

$$\begin{aligned} F(x) &= \int_{-\infty}^x f(z) dz, \\ f(x) &= F'(x). \end{aligned}$$

### Eksempel 2.2. Fordelingsfunktionen for den geometriske fordeling

Lad  $X$  følge den geometriske fordeling med parameter  $p$ , jf. Eksempel 2.1. Sandsynlighedsfunktionen for  $X$  er altså  $f(x) = p(1-p)^x, x = 0, 1, 2, \dots$ . Vi vil finde den kumulerede fordelingsfunktion  $F$ .

For det første er  $F(x) = 0$  når  $x < 0$ , og hvis  $x$  er et ikke-negativt heltal er

$$\begin{aligned} F(x) &= \sum_{z \leq x} f(z) \\ &= \sum_{z=0}^x p(1-p)^z \\ &= p \frac{1 - (1-p)^{x+1}}{1 - (1-p)} \\ &= 1 - (1-p)^{x+1} \end{aligned}$$

ifølge reglen om summation af en kvotientrække.

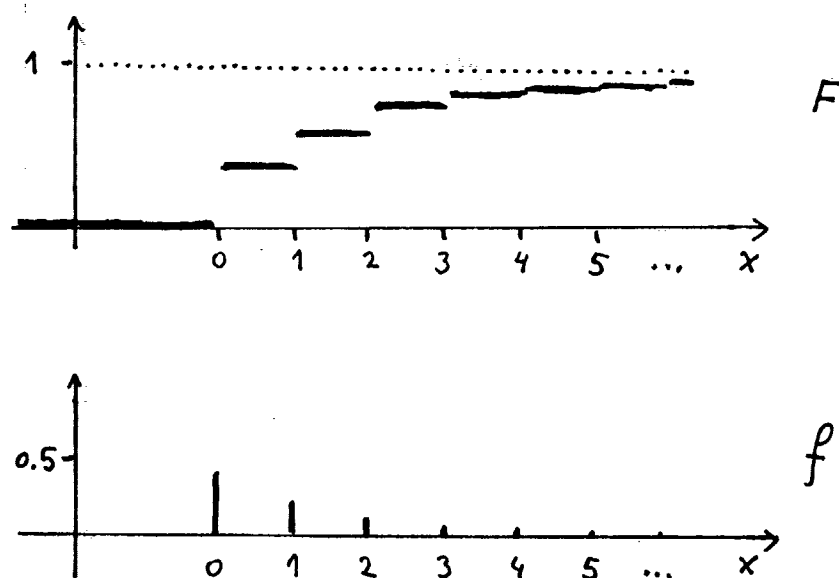
Omvendt er

$$\begin{aligned} F(x) - F(x-1) &= (1-p)^x - (1-p)^{x+1} \\ &= p(1-p)^x \\ &= f(x), \end{aligned}$$

hvilket illustrerer (2.4).

I Figur 2.6 illustreres sammenhængen mellem  $F$  og  $f$ .  $f(x)$  er  $F$ 's spring i punktet  $x$ , og  $F(x)$  er summen af  $f$ -værdierne til venstre for eller i  $x$ .  $\square$

<sup>1</sup>I differential- og integralregningens forstand er  $F$  en stamfunktion til  $f$  og  $f$  er den afledede af  $F$ .



**Figur 2.6:** Fordelingsfunktionen  $F$  og sandsynlighedsfunktionen  $f$  for den geometriske fordeling med  $p = 0.382$ .

**Eksempel 2.3. Eksponentialfordelingen**

Den geometriske fordeling blev indført i Eksempel 2.1 som fordelingen af "ventetiden" indtil en bestemt begivenhed (mønten viser Krone) indtræffer for første gang. Denne "ventetid" måles som "antal gange et bestemt forsøg skal gentages indtil den ønskede begivenhed indtræffer".

Nu er man tit interesseret i at opstille sandsynlighedsmodeller for rigtige ventetider, altså sådan nogen som angives på den sædvanlige kontinuerte tidsskala og som måles ved hjælp af et ur. Man kunne eksempelvis søge en fornuftig sandsynlighedsmodel for ventetiden indtil en bestemt slags begivenhed indtræffer, forudsat at begivenhederne indtræffer "fuldstændig tilfældigt". I Basisstatistik-noterne bliver der argumenteret for, at sådanne ventetider undertiden kan beskrives med en eksponentialfordeling.

Eksponentialfordelingen med parameter  $\lambda$  er den kontinuerte sandsynlighedsfordeling (på udfaldsrummet  $\mathcal{X} = ]0, +\infty[$ ) med sandsynlighedsfunktion

$$f(x) = \lambda \exp(-\lambda x), \quad x > 0.$$

Her er  $\lambda$  en positiv parameter som beskriver den intensitet hvormed begivenhederne indtræffer (dvs. desto større  $\lambda$ , jo oftere optræder begivenhederne).

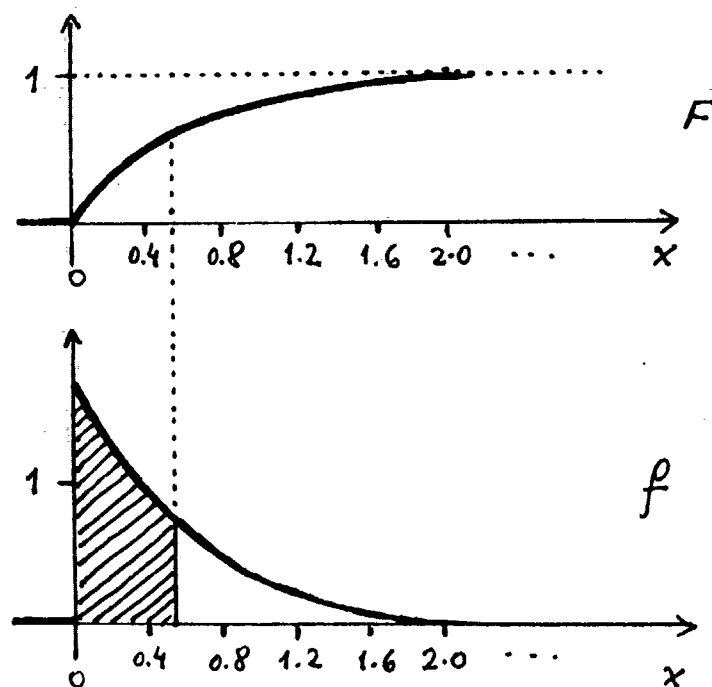
Den kumulerede fordelingsfunktion for eksponentialfordelingen er (for  $x > 0$ )

$$\begin{aligned} F(x) &= \int_{-\infty}^x f(z) dz \\ &= \int_0^x \lambda \exp(-\lambda z) dz \\ &= 1 - \exp(-\lambda x). \end{aligned}$$

I Figur 2.7 illustreres sammenhængen mellem  $F$  og  $f$ .  $f(x)$  er differentialkvotienten af  $F$  i punktet  $x$ , og  $F(x)$  er arealet under kurven  $f$  på intervallet  $] -\infty, x]$ .  $\square$

**2.4 Fraktiler**

Fordelingsfunktionen  $F$  for en sandsynlighedsfordeling er løsningen på alle opgaverne



Figur 2.7: Fordelingsfunktionen  $F$  og sandsynlighedstæthedsfunktionen  $f$  for eksponentialfordelingen med  $\lambda = 1.618$ .

Man har et bestemt  $x$ . Man søger sandsynligheden for at få en observation til venstre for  $x$ .

I visse forbindelser, bl.a. i forbindelse med test af statistiske hypoteser, har man brug for at løse de omvendte opgaver

Man har et tal  $\alpha$  mellem 0 og 1. Man søger et  $x_\alpha$  med den egenskab, at sandsynligheden er  $\alpha$  for at få en observation til venstre for  $x_\alpha$ .

Et sådant tal  $x_\alpha$  kaldes en  $\alpha$ -fraktil i den pågældende sandsynlighedsfordeling. Opgaven at bestemme  $\alpha$ -fraktiler kan lidt mere formelt formuleres som opgaven at løse ligningen

$$F(x) = \alpha \quad (2.5)$$

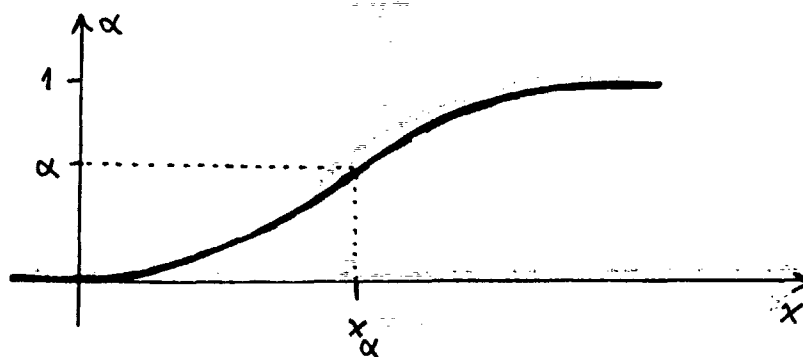
med hensyn til  $x$ , og problemerne opstår derved at fordelingsfunktionen  $F$  ikke nødvendigvis er bijektiv. Det eneste man kan udlede af definitionen (2.3) er, at  $F$  er voksende og kontinuert fra højre, og det medfører ikke at (2.5) altid har præcis en løsning. Figurerne 2.8-2.10 illustrerer hvad slags problemer man kommer ud for. Som det fremgår af figurerne er det ikke helt oplagt hvordan man skal definere en  $\alpha$ -fraktil. Man plejer at benytte følgende definition:

En  $\alpha$ -fraktil for sandsynlighedsfordelingen med fordelingsfunktion  $F$  er et tal  $x_\alpha$  med den egenskab, at  $F(x_\alpha - h) \leq \alpha$  og  $F(x_\alpha + h) \geq \alpha$  for alle  $h > 0$ .

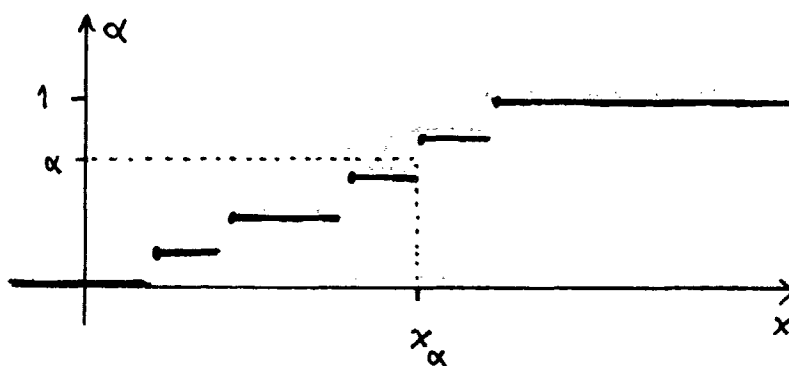
Efter denne definition findes der til hvert  $\alpha \in ]0, 1[$  mindst én  $\alpha$ -fraktil, og mængden af  $\alpha$ -fraktiler udgør et afsluttet interval, eventuelt et enkelt punkt.

## 2.5 Flerdimensionale fordelinger

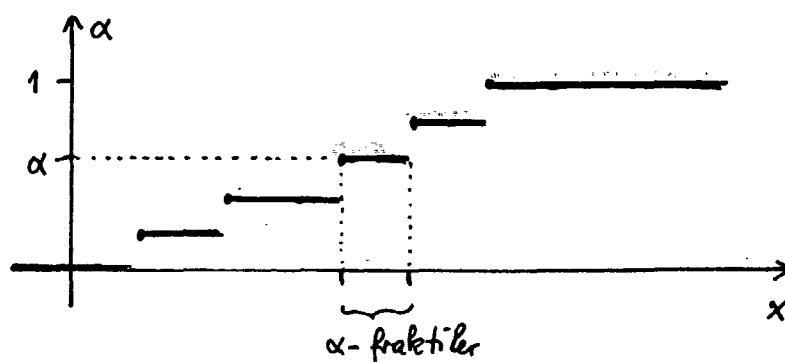
Også  $n$ -dimensionale stokastiske variable kan have en kontinuert sandsynlighedsfordeling. Lad der være tale om en  $n$ -dimensional stokastisk variabel  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  på udfaldsrummet  $\mathcal{X} = \mathbf{R}^n$  (eller en



Figur 2.8: – her er én  $\alpha$ -fraktil  $x_\alpha$ , og  $F(x_\alpha) = \alpha$ .

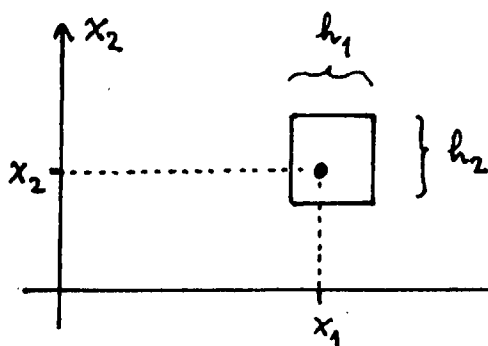


Figur 2.9: – her er også én  $\alpha$ -fraktil  $x_\alpha$ , men  $F(x_\alpha) \neq \alpha$ .



Figur 2.10: – her er et helt interval af  $\alpha$ -fraktiler.  $F(x_\alpha) = \alpha$  for dem alle med undtagelse af intervallets højre endepunkt.





Figur 2.11: Et  $h$ -interval omkring  $\mathbf{x} = (x_1, x_2)$ .

passende delmængde heraf). I lighed med det just behandlede endimensionale tilfælde antages det, at sandsynlighedsfordelingen for  $\mathbf{X}$  kan beskrives ved en sandsynlighedstæthedsfunktion

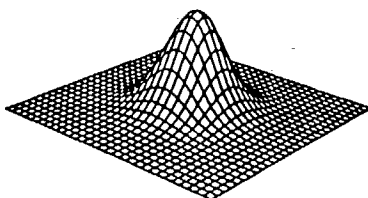
$$f(\mathbf{x}) = \lim_{h \searrow 0} \frac{P(\mathbf{X} \text{ ligger i et } h\text{-interval omkring } \mathbf{x})}{\text{"størrelsen" af } h}.$$

Her betegner  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  et punkt i udfaldsrummet, og  $\mathbf{h} = (h_1, h_2, \dots, h_n)$  er en  $n$ -tupel af positive tal. Endvidere er "et  $h$ -interval omkring  $\mathbf{x}$ " en mængde af formen  $I_1 \times I_2 \times \dots \times I_n$  hvor  $I_j$  er et  $h_j$ -interval omkring  $x_j$ , og endelig er "størrelsen" af  $\mathbf{h}$  ganske enkelt tallet  $h_1 h_2 \dots h_n$ .

For at vise hvad meningen er, ser vi på det to-dimensionale tilfælde, hvor vi kan lave tegninger. I Figur 2.11 illustreres et  $h$ -interval: det er et rektangel, og størrelsen af  $\mathbf{h}$  (eller af  $h$ -intervallet) er ikke andet end arealet af rektanglet. Figur 2.12 er et forsøg på at vise, hvordan en sandsynlighedstæthedsfunktion for en to-dimensional stokastisk variabel  $\mathbf{X} = (X_1, X_2)$  kan se ud.

I det en-dimensionale tilfælde kunne man bestemme sandsynligheder ved at regne integraler ud. Det kan man også i det  $n$ -dimensionale tilfælde, blot bliver det her  $n$ -dobbelte integraler. Hvis vi forsigtigt holder os til det to-dimensionale tilfælde er f.eks.

$$\begin{aligned} &P(a_1 < X_1 \leq b_1 \text{ og } a_2 < X_2 \leq b_2) \\ &= \int_{a_1}^{b_1} \int_{a_2}^{b_2} f(x_1, x_2) dx_2 dx_1. \end{aligned} \quad (2.6)$$



**Figur 2.12:** Tæthedsfunktion for en todimensional fordeling.

Også  $n$ -dimensionale tætheder skal (være ikke-negative og) integrere til 1, for  $n = 2$  altså specielt

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x_1, x_2) dx_2 dx_1 = 1.$$

Når man kender den simultane tæthedsfunktion for nogle stokastiske variable, kan man let bestemme sandsynligheder og tæthedsfunktioner for et eller andet delsæt af de variable. Antag f.eks. at  $(X_1, X_2)$  har en simultan tæthedsfunktion  $f(x_1, x_2)$ . Så får man ved at benytte (2.6)

$$\begin{aligned} P(a < X_1 \leq b) &= P(a < X_1 \leq b \text{ og } -\infty < X_2 < +\infty) \\ &= \int_a^b \int_{-\infty}^{+\infty} f(x_1, x_2) dx_2 dx_1 \\ &= \int_a^b \left( \int_{-\infty}^{+\infty} f(x_1, x_2) dx_2 \right) dx_1. \end{aligned}$$

Heraf fremgår det, at funktionen

$$f_1(x) = \int_{-\infty}^{+\infty} f(x, x_2) dx_2 \quad (2.7)$$

er tæthedsfunktion for  $X_1$  i dennes marginale fordeling.

Definitionen på stokastisk uafhængighed (side 16) gælder helt uændret for flerdimensionale kontinuerte variable. Desuden gælder en pendant til Sætning 1.1:

**Sætning 2.1** *Kontinuerte stokastiske variable  $X_1, X_2, \dots, X_n$  er stokastisk uafhængige, hvis og kun hvis de har en simultan sandsynlighedstæthedsfunktion  $f(x_1, x_2, \dots, x_n)$  som kan skrives som et produkt af de enkelte komponenters marginale sandsynlighedstæthedsfunktioner, altså som*

$$f(x_1, x_2, \dots, x_n) = f_1(x_1)f_2(x_2) \dots f_n(x_n)$$

hvor  $f_i(x_i)$  er den marginale tæthedsfunktion for  $X_i$ .

## 2.6 Transformation af kontinuerte fordelinger

I afsnit 1.2 har vi beskæftiget os med operationen transformation af en sandsynlighedsfordeling og bestemmelse af betingede sandsynligheder. Disse tidligere overvejelser gælder helt uden videre for diskrete sandsynlighedsfordelinger på den reelle akse, hvorimod der for de kontinuerte fordelingers vedkommende er forskellige tekniske modifikationer. Men først kan der være grund til at dvæle et øjeblik ved den diskrete udgave af en bestemt type transformationer af stokastiske variable, nemlig *summer* af stokastiske variable.

**Eksempel 2.4.** *Eksempel på bestemmelse af fordelingen af summen  $Y = X_1 + X_2$  af to heltallige stokastiske variable  $X_1$  og  $X_2$*

Antag at  $X_1$  og  $X_2$  er heltallige stokastiske variable med simultan sandsynlighedsfunktion

$$f(x_1, x_2) = P(X_1 = x_1, X_2 = x_2).$$

Lad os sige at udfaldsrummet er  $\mathcal{X} = \{1, 2, 3\} \times \{1, 2, 3\}$ , hvilket kan illustreres således:

3	•	•	•
2	•	•	•
1	•	•	•
	1	2	3

På næste illustration er angivet de delmængder af udfaldsrummet der svarer til at  $y = x_1 + x_2$  er noget bestemt.

3			
2			
1			
	1	2	3

Eksempelvis er den del af udfaldsrummet der svarer til  $y = 3$  punkterne (2,1) og (1,2); sandsynligheden for at  $Y$  er lig 3 er derfor summen af sandsynlighederne for at  $(X_1, X_2)$  er (2,1) og for at  $(X_1, X_2)$  er (1,2).

Hvis den simultane fordeling af  $X_1$  og  $X_2$  f.eks. er givet ved

3	0.2	0.1	0.1
2	0.1	0.1	0.1
1	0.1	0.1	0.1
	1	2	3

så ser fordelingen af  $Y = X_1 + X_2$  således ud:

$y$	$P(Y = y)$
2	0.1
3	0.2
4	0.4
5	0.2
6	0.1
	1

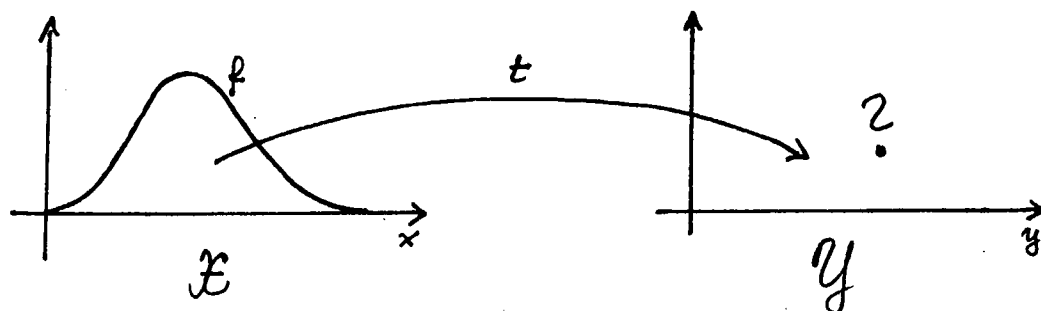
Den stokastiske variabel  $Y = X_1 + X_2$  har da (ifølge den generelle regel (1.1) på side 7) sandsynlighedsfunktionen

$$\begin{aligned}
 g(y) &= \sum_{(x_1, x_2) : x_1 + x_2 = y} f(x_1, x_2) \\
 &= \sum_x f(y - x, x) .
 \end{aligned}$$

□

Generelt gælder, at hvis  $X_1$  og  $X_2$  er stokastisk uafhængige heltallige variable, så er (ifølge Sætning 1.1) deres simultane sandsynlighedsfunktion et produkt af de marginale sandsynlighedsfunktioner for  $X_1$  og  $X_2$ ,

$$f(x_1, x_2) = f_1(x_1) \times f_2(x_2) ,$$



Figur 2.13: Opgave: Transformation af en tæthed.

og formelen for sandsynlighedsfunktionen for  $Y = X_1 + X_2$  bliver da

$$g(y) = \sum_x f_1(y-x)f_2(x). \quad (2.8)$$

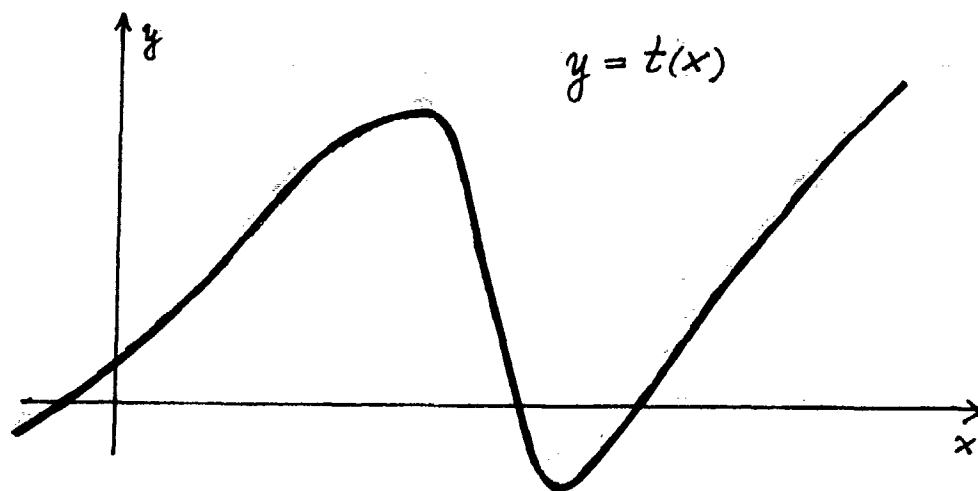
Vi skal i næste afsnit se hvordan man bestemmer tæthedsfunktionen for en sum af to *kontinuerte* stokastiske variable.

Vi vil nu beskæftige os med opgaven transformation af en endimensional kontinuert sandsynlighedsfordeling, dvs. opgaven: bestem tæthedsfunktionen  $g(y)$  for den transformerede stokastiske variabel  $Y = t(X)$ , når  $X$  har tæthedsfunktionen  $f(x)$ , og når afbildningen  $t$  er tilstrækkelig "pæn". Man kan også kalde opgaven for *transformation af tætheder*. Figur 2.13 illustrerer opgaven.

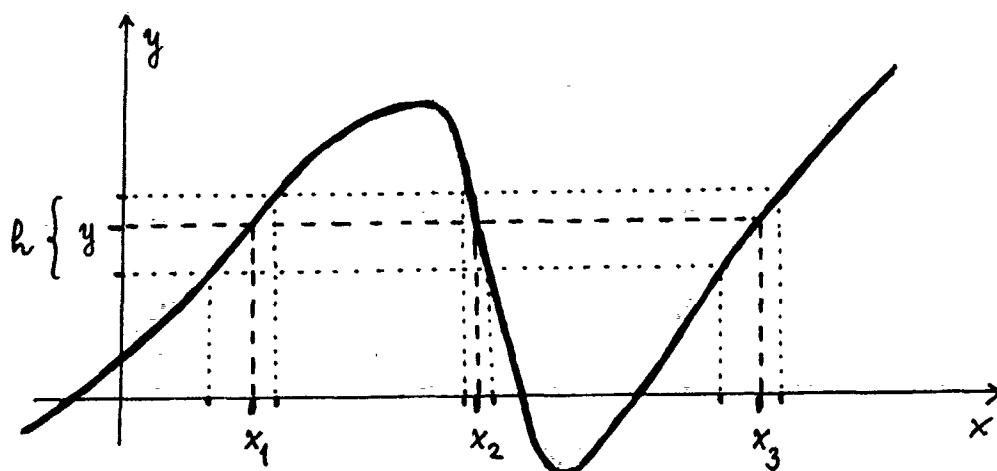
Da det går ud på at bestemme  $g(y)$  der skal være

$$\lim_{h \searrow 0} \frac{P(Y \text{ ligger i et } h\text{-interval omkring } y)}{h},$$

ser vi først på  $P(Y \text{ ligger i et } h\text{-interval omkring } y)$ . Det hele afhænger af hvordan transformationen  $t$  nærmere er indrettet, men antag f.eks. at  $t$  er som skitseret på Figur 2.14. Nu indtegnes det punkt  $y$  i hvilket vi skal finde  $g(y)$ , samt et  $h$ -interval omkring  $y$ , og vi får Figur 2.15. (På figurerne er  $t$  tegnet pæn og glat; vi vil faktisk til de følgende overvejelser antage, at  $t$  er kontinuert differentiabel og at differentialkvotienten  $t'$  er forskellig fra 0 undtagen i et endeligt antal punkter.)



Figur 2.14: Afbildningen  $t$ .



Figur 2.15: Bestemmelse af de  $x$ -intervaller som ved  $t$  afbildes over i et  $h$ -interval omkring  $y$ .

Det ses at der er tre punkter  $x_1, x_2$  og  $x_3$  som ved  $t$  afbildes over i  $y$ . Omkring hvert af disse  $x$ -er er der et interval der afbildes over i  $h$ -intervallet omkring  $y$ , og disse intervaller på  $x$ -aksen er desto snævrere jo stejlere grafen er. Faktisk er længden af  $x$ -intervallet nogenlunde lig med længden af  $y$ -intervallet (som er  $h$ ) divideret med grafens hældningskoefficient i det pågældende  $x$ : længden af intervallet omkring  $x_i$  er ca.  $h/|t'(x_i)|$ . Derfor er

$$\begin{aligned} & P(Y \text{ ligger i } h\text{-intervallet omkring } y) \\ & \approx P(X \text{ ligger i et } \frac{h}{|t'(x_1)|}\text{-interval omkring } x_1) + \\ & \quad P(X \text{ ligger i et } \frac{h}{|t'(x_2)|}\text{-interval omkring } x_2) + \\ & \quad P(X \text{ ligger i et } \frac{h}{|t'(x_3)|}\text{-interval omkring } x_3) \\ & \approx f(x_1) \frac{h}{|t'(x_1)|} + f(x_2) \frac{h}{|t'(x_2)|} + f(x_3) \frac{h}{|t'(x_3)|}. \end{aligned}$$

Ved nu at dividere med  $h$  og lade  $h$  gå mod 0 får vi

$$g(y) = \frac{f(x_1)}{|t'(x_1)|} + \frac{f(x_2)}{|t'(x_2)|} + \frac{f(x_3)}{|t'(x_3)|}.$$

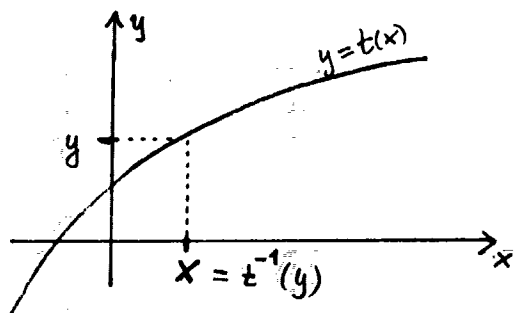
Dette eksempel skulle forhåbentlig vise, hvordan man i al almindelighed kan finde tætheden  $g(y)$  for  $Y = t(X)$  når man kender tætheden  $f(x)$  for  $X$ . I mange forekommende tilfælde er afbildningen  $t$  strengt voksende, således at der til et  $y$  (højst) findes ét  $x$  så  $t(x) = y$ . I så fald er opskriften til bestemmelse af  $g(y)$  den, at

$$g(y) = f(x)/|t'(x)|$$

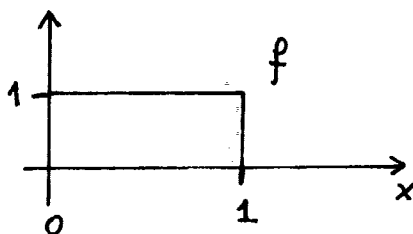
når  $t(x) = y$ , dvs. når  $x = x(y) = t^{-1}(y)$  ( $t^{-1}$  betegner den omvendte afbildning til  $t$ ), jf. Figur 2.16. Da  $1/t'(t^{-1}(y))$  er lig med  $(t^{-1})'(y)$ , kan opskriften på  $g(y)$  skrives som

$$g(y) = f(t^{-1}(y)) \times |(t^{-1})'(y)|. \quad (2.9)$$

Man kan her tænke på  $|(t^{-1})'(y)|$  som det reciprokke forhold mellem længden af et lille  $h$ -interval omkring  $y$  og længden af det dertil svarende interval omkring  $x = t^{-1}(y)$ .



Figur 2.16:  $x = x(y) = t^{-1}(y)$ .



Figur 2.17: Tætheden for ligefordelingen på  $]0, 1]$ .

**Eksempel 2.5. Transformation af ligefordeling til eksponentialfordeling. Simulation af eksponentialfordelte variable**

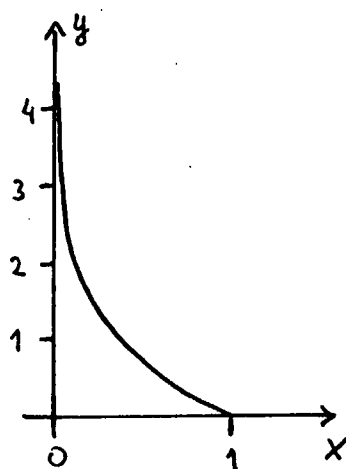
Antag at  $X$  er ligefordelt på intervallet  $]0, 1]$ , dvs. at  $X$  har sandsynlighedstætheden

$$f(x) = \begin{cases} 1 & \text{hvis } x \in ]0, 1] \\ 0 & \text{ellers,} \end{cases}$$

se Figur 2.17.

Man får nu den besynderlige idé at ville transformere  $X$  med afbildningen  $t : x \mapsto -\frac{1}{\lambda} \ln x$ , hvor  $\lambda$  er en positiv konstant.  $t$  afbilder intervallet  $]0, 1]$  bijektivt på  $[0, +\infty[$ , se Figur 2.18. For at bestemme den omvendte afbildning  $t^{-1}$  løses ligningen  $y = t(x)$ , dvs.  $y = -\frac{1}{\lambda} \ln x$ , mht.  $x$ , hvilket giver  $x = \exp(-\lambda y)$ , dvs.  $t^{-1}(y) = \exp(-\lambda y)$ . Så er  $(t^{-1})'(y) = -\lambda \exp(-\lambda y)$ , og





Figur 2.18: Grafen for  $t : x \mapsto -\frac{1}{\lambda} \ln x$ .

ifølge (2.9) er tætheden for den transformerede stokastiske variabel  $Y = t(X)$

$$\begin{aligned} g(y) &= f(\exp(-\lambda y)) \times |-\lambda \exp(-\lambda y)| \\ &= \lambda \exp(-\lambda y), \quad y \geq 0, \end{aligned}$$

fordi når  $y \geq 0$ , er  $\exp(-\lambda y) \in ]0, 1]$ , og dermed er  $f(\exp(-\lambda y))$  lig med 1.

Det viser sig således, at  $Y$  er eksponentialfordelt med parameter  $\lambda$ .

Dette resultat er interessant af følgende grund: Mange lommeregner og datamater har en anordning som kan frembringe (pseudo-)tilfældige tal fra ligefordelingen på  $]0, 1]$ . Ved at transformere sådanne tilfældige tal med afbildningen  $x \mapsto -\frac{1}{\lambda} \ln x$  får man ifølge det netop udledte tilfældige tal fra eksponentialfordelingen med parameter  $\lambda$ . Vi har altså en simpel metode til at simulere eksponentialfordelte tal.  $\square$

## 2.7 Transformation af flerdimensionale variable

Man har tit brug for også at kunne transformere flerdimensionale stokastiske variable. Det bliver hurtigt noget teknisk, men det skal dog

nævnes, at i det "pæne" tilfælde hvor  $t$  er en en-entydig afbildning således at både  $t$  og  $t^{-1}$  er kontinuert differentiable, så gælder "den samme" formel

$$g(\mathbf{y}) = f(t^{-1}(\mathbf{y})) \times |(t^{-1})'(\mathbf{y})| \quad (2.10)$$

for tætheden  $g$  for  $\mathbf{Y} = t(\mathbf{X})$  udtrykt ved tætheden  $f$  for  $\mathbf{X}$ . Vi mangler dog at fortælle hvad  $|(t^{-1})'(\mathbf{y})|$  betyder i dette tilfælde: man skal i denne omgang tænke på tallet  $|(t^{-1})'(\mathbf{y})|$  som det reciprokke forhold mellem "volumenet" af et lille  $h$ -interval omkring  $\mathbf{y}$  og "volumenet" af den tilsvarende omegn omkring  $\mathbf{x}$ . Mere præcist er  $|(t^{-1})'(\mathbf{y})|$  den såkaldte *funktionaldeterminant*

$$\left| \det \begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \dots & \frac{\partial x_1}{\partial y_n} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \dots & \frac{\partial x_2}{\partial y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial y_1} & \frac{\partial x_n}{\partial y_2} & \dots & \frac{\partial x_n}{\partial y_n} \end{pmatrix} \right|$$

udregnet i punktet  $\mathbf{y} = (y_1, y_2, \dots, y_n)$ .

### Eksempel 2.6. En sum af to eksponentialfordelte variable

Antag at  $X_1$  og  $X_2$  er uafhængige identisk eksponentialfordelte stokastiske variable med parameter  $\lambda > 0$ . Hvad er fordelingen af  $X_1 + X_2$ ?

Denne opgave kan løses således. Sæt

$$Y_1 = X_1 + X_2,$$

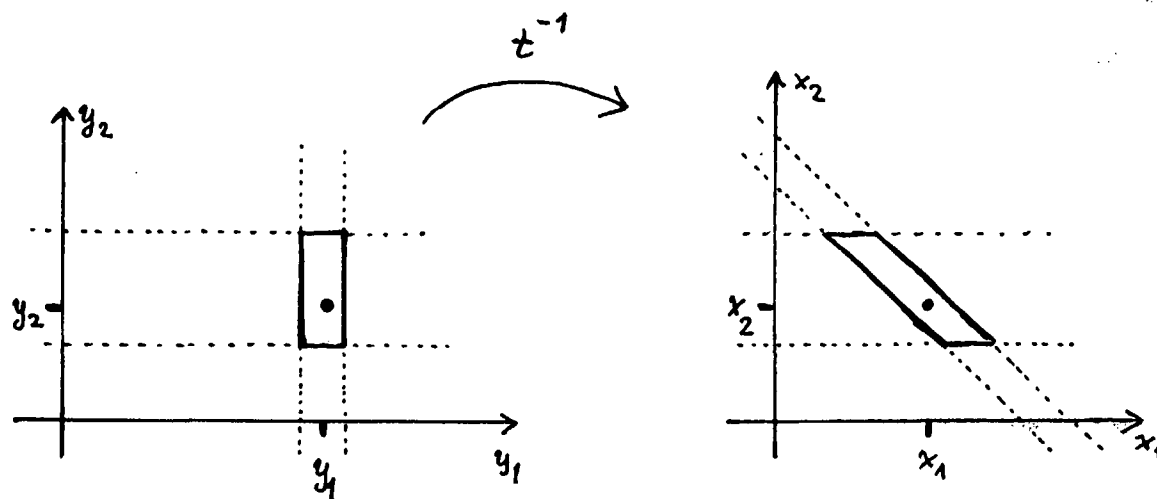
$$Y_2 = X_2.$$

Så er  $\mathbf{Y} = (X_1, X_2)$  en funktion  $t$  af  $\mathbf{X} = (X_1, X_2)$ , og vi kan finde fordelingen af  $\mathbf{Y}$ . Når vi har fundet den, kan vi bagefter finde fordelingen af  $Y_1$  som er den vi egentlig er interesserede i.

Ligningssystemet

$$y_1 = x_1 + x_2$$

$$y_2 = x_2$$



Figur 2.19: Afbildningen  $t^{-1}$  i Eksempel 2.6.

er ensbetydende med

$$x_1 = y_1 - y_2$$

$$x_2 = y_2$$

Det betyder at  $t^{-1}$  afbilder  $(y_1, y_2)$  over i  $(x_1, x_2) = (y_1 - y_2, y_2)$ . Man kan nu sætte sig til at regne ud, hvordan et lille  $h$ -interval omkring  $\mathbf{y}$  afbildes ved  $t^{-1}$ , se Figur 2.19. Det viser sig, at et rektangel omkring  $\mathbf{y} = (y_1, y_2)$  afbildes over i et parallelogram omkring  $\mathbf{x} = (x_1, x_2) = t^{-1}(\mathbf{y})$ . Parallelogrammet har samme grundflade og samme højde som rektanglet, og derfor bliver "volumen-forholdet"  $|(t^{-1})'(\mathbf{y})|$  lig 1.

(Det kunne man også finde ved at udregne funktionaldeterminanten: For det første er

$$\begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix},$$

og dernæst er  $\det \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} = 1 \times 1 - 0 \times (-1) = 1$ .)

Ifølge formelen (2.10) for transformation af tætheder er tætheden for  $(Y_1, Y_2)$

$$\begin{aligned} g(y_1, y_2) &= f(t^{-1}(x_1, x_2)) \\ &= f(y_1 - y_2, y_2). \end{aligned}$$

Da  $X_1$  og  $X_2$  er stokastisk uafhængige identisk eksponentialfordelte med parameter  $\lambda > 0$ , er

$$\begin{aligned} f(x_1, x_2) &= \lambda \exp(-\lambda x_1) \times \lambda \exp(-\lambda x_2) \\ &= \lambda^2 \exp(-\lambda(x_1 + x_2)) \end{aligned}$$

når  $x_1 > 0$  og  $x_2 > 0$ . Dermed er

$$g(y_1, y_2) = \lambda^2 \exp(-\lambda y_1).$$

Dette gælder for de  $(y_1, y_2)$  der ligger i billedmængden for  $t$ , dvs. de  $(y_1, y_2)$  der kan skrives som  $t(x_1, x_2)$  for  $x_1 > 0$  og  $x_2 > 0$ , og det er det samme som de  $(y_1, y_2)$  for hvilke  $y_1 - y_2 > 0$  og  $y_2 > 0$ , dvs.  $y_1 > y_2 > 0$ . Tætheden for  $(Y_1, Y_2)$  er altså

$$g(y_1, y_2) = \begin{cases} \lambda^2 \exp(-\lambda y_1) & \text{når } y_1 > y_2 > 0, \\ 0 & \text{ellers.} \end{cases}$$

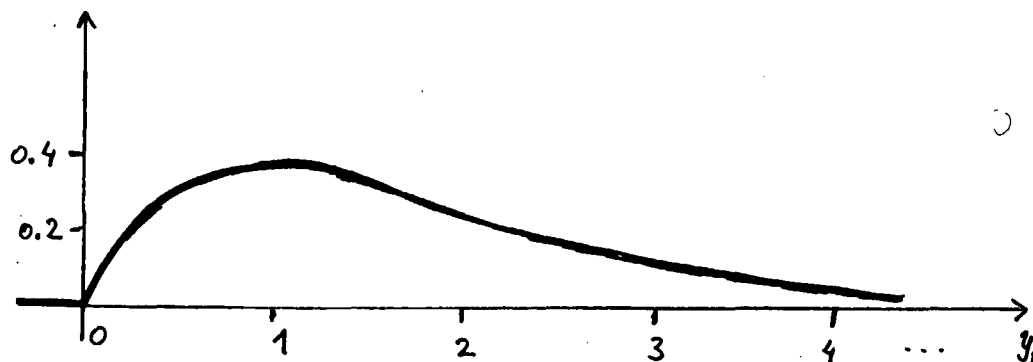
Den marginale tæthed for  $Y_1$  er ifølge (2.7)

$$\begin{aligned} g_1(y_1) &= \int_{-\infty}^{+\infty} g(y_1, y_2) dy_2 \\ &= \int_0^{y_1} \lambda^2 \exp(-\lambda y_1) dy_2 \\ &= \lambda^2 y_1 \exp(-\lambda y_1) \end{aligned} \tag{2.11}$$

når  $y_1 > 0$ , jf. Figur 2.20.

Vi har hermed bestemt sandsynlighedstæthedsfunktionen for (fordelingen af) summen af to uafhængige identisk eksponentialfordelte stokastiske variable<sup>2</sup>.  $\square$

<sup>2</sup>Det kan i øvrigt nævnes, at den fundne fordeling er en såkaldt *gammafordeling* med formparameter 2 og skalaparameter  $\lambda^{-1}$ . Hvis specielt  $\lambda = 1/2$ , så hedder fordelingen også en  $\chi^2$ -fordeling med 4 frihedsgrader.



Figur 2.20: Tætheden for  $Y_1$  i Eksempel 2.6.

I ovenstående eksempel fandt vi fordelingen af en sum af to eksponentialfordelte stokastiske variable. Selve fremgangsmåden beroede ikke på at det specielt var eksponentialfordelte stokastiske variable, og man kan da også benytte metoden generelt. Man vil da finde, at hvis  $(X_1, X_2)$  har simultan tæthed  $f(x_1, x_2)$ , så er tætheden  $g(y)$  for  $Y = X_1 + X_2$

$$g(y) = \int_{-\infty}^{+\infty} f(y-x, x) dx.$$

Den tilsvarende formel for sandsynlighedsfunktioner for diskrete stokastiske variable er (jf. (2.8))

$$g(y) = \sum_x f(y-x, x).$$

Hvis specielt  $X_1$  og  $X_2$  er stokastisk uafhængige, så har man formlerne

$$g(y) = \int_{-\infty}^{+\infty} f_1(y-x) f_2(x) dx \quad (2.12)$$

og

$$g(y) = \sum_x f_1(y-x) f_2(x) \quad (2.13)$$

for sandsynlighedstæthedsfunktionen hhv. sandsynlighedsfunktion for en sum af to hhv. kontinuerte og diskrete variable.

Som afslutning på gennemgangen af transformation af sandsynlighedsfordelinger/stokastiske variable et lidt mere sofistikeret eksempel, der benyttes i Kapitel 4.

### Eksempel 2.7. Transformation ved hjælp af polære koordinater

Antag at  $(X_1, X_2)$  følger en eller anden to-dimensional kontinuert fordeling med tæthed  $f(x_1, x_2)$ . Man kan tænke på  $(X_1, X_2)$  som et punkt i planen, og man vil nu angive punktet i en slags polære koordinater idet man angiver

- den kvadratiske afstand  $Y_1$  mellem  $(0, 0)$  og  $(X_1, X_2)$ , og
- den vinkel  $Y_2$  som linien fra  $(0, 0)$  gennem  $(X_1, X_2)$  danner med abscisseaksen,

se Figur 2.21. Derfor betragtes afbildningen  $(y_1, y_2) = t(x_1, x_2)$  givet ved

$$\begin{aligned} y_1 &= x_1^2 + x_2^2 \\ y_2 &= \begin{cases} \cos^{-1}(x_1/y_1^{1/2}) & \text{hvis } x_2 \geq 0, \\ 2\pi - \cos^{-1}(x_1/y_1^{1/2}) & \text{hvis } x_2 < 0. \end{cases} \end{aligned}$$

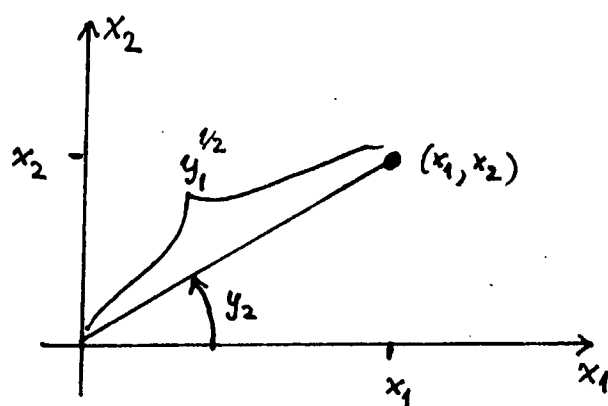
(Denne afbildning  $t$  afbilder  $\mathbf{R}^2 \setminus \{(0, 0)\}$  bijektivt på  $]0, +\infty[ \times ]0, 2\pi[$ ). Opgaven er da at bestemme tætheden for  $(Y_1, Y_2) = t(X_1, X_2)$  udtrykt ved tætheden  $f(x_1, x_2)$  for  $(X_1, X_2)$ .

Den omvendte afbildning til  $t$  er  $(x_1, x_2) = t^{-1}(y_1, y_2)$  givet ved

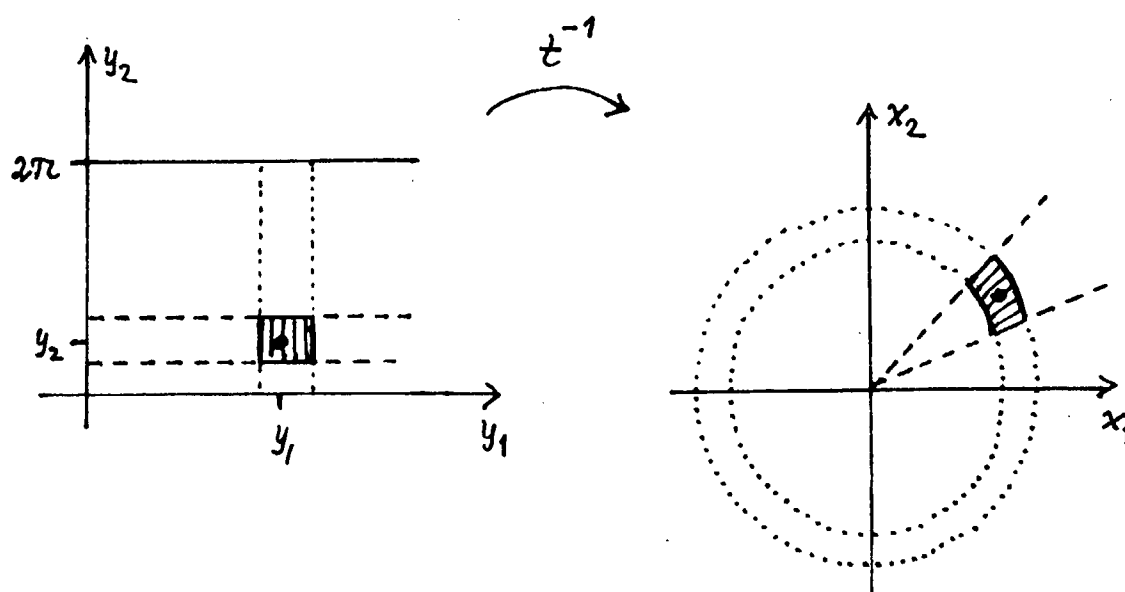
$$\begin{aligned} x_1 &= y_1^{1/2} \cos y_2 \\ x_2 &= y_1^{1/2} \sin y_2. \end{aligned}$$

Man kan nu undersøge hvordan billedet af et  $h$ -interval omkring  $\mathbf{y} = (y_1, y_2)$  afbildes ved  $t^{-1}$ . Læseren opfordres til at følge med på Figur 2.22. Lad os sige at  $h$ -intervallet omkring  $\mathbf{y}$  er givet som

$$]y_1 - \tfrac{1}{2}h_1, y_1 + \tfrac{1}{2}h_1[ \times ]y_2 - \tfrac{1}{2}h_2, y_2 + \tfrac{1}{2}h_2[.$$



Figur 2.21: Eksempel 2.7:  $(y_1, y_2)$  bestemt ud fra  $(x_1, x_2)$ .



Figur 2.22: Eksempel 2.7: Hvordan  $t^{-1}$  transformerer et  $h$ -interval.

Ved  $t^{-1}$  afbildes dette over i et område som ligger dels mellem de to cirkler med radius hhv.  $(y_1 - \frac{1}{2}h_1)^{1/2}$  og  $(y_1 + \frac{1}{2}h_1)^{1/2}$ , dels mellem to radier hvis vinkelafstand er  $h_2$ . Arealet af billedet af  $h$ -intervallet er

$$\frac{h_2}{2\pi} \times \left( \pi(y_1 + \frac{1}{2}h_1) - \pi(y_1 - \frac{1}{2}h_1) \right) = \frac{1}{2}h_1h_2,$$

dvs. "volumenforholdet"  $|(t^{-1})'(y)|$  er  $1/2$ .

Vi kunne finde det samme ved at udregne funktionaldeterminanten:

$$\begin{aligned} \left| \begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{pmatrix} \right| &= \left| \begin{pmatrix} \frac{1}{2}y_1^{-1/2} \cos y_2 & -y_1^{1/2} \sin y_2 \\ \frac{1}{2}y_1^{-1/2} \sin y_2 & y_1^{1/2} \cos y_2 \end{pmatrix} \right| \\ &= \frac{1}{2}(\cos y_1)^2 + \frac{1}{2}(\sin y_1)^2 \\ &= \frac{1}{2}. \end{aligned}$$

Ifølge formelen (2.10) for transformation af tætheder er tætheden for  $(Y_1, Y_2)$

$$g(y_1, y_2) = \frac{1}{2}f(y_1^{1/2} \cos y_2, y_1^{1/2} \sin y_2).$$

□

## 2.8 Betingede fordelinger

Herefter går vi over til opgaven at bestemme betingede sandsynligheder når talen er om kontinuerte sandsynlighedsfordelinger.

Se igen på "eksemplet" fra side 39-41, og lad os nu finde den betingede fordeling af  $X$  givet at  $t(X) = y$ . Det er klart at denne betingede fordeling må være beliggende på  $\mathcal{X}_y = \{x : t(x) = y\} = \{x_1, x_2, x_3\}$ , men hvordan skal sandsynlighedsmassen fordeles imellem disse tre punkter? For at afgøre dette udvider vi den betingende hændelse og betinger med at  $t(X)$  ligger i et af de tre småintervaller på  $x$ -aksen, og f.eks. er da (jf. Figur 2.15)

$$\begin{aligned} &P(X \text{ ligger i interval nr. 1} \mid t(X) \text{ ligger i } h\text{-intervallet om } y) \\ &= \frac{P(X \text{ ligger i interval nr. 1})}{P(t(X) \text{ ligger i } h\text{-intervallet om } y)} \end{aligned}$$



$$\begin{aligned}
&= \frac{P(X \text{ ligger i interval nr. 1})}{P(X \text{ ligger i et af de tre intervaller})} \\
&\approx \frac{f(x_1) \times h/|t'(x_1)|}{f(x_1) \times h/|t'(x_1)| + f(x_2) \times h/|t'(x_2)| + f(x_3) \times h/|t'(x_3)|} \\
&= \frac{f(x_1)/|t'(x_1)|}{g(y)},
\end{aligned}$$

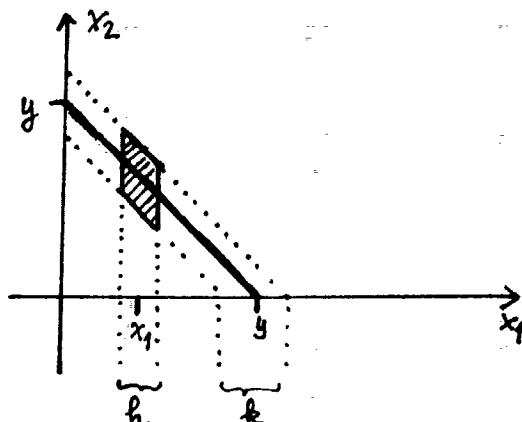
hvor  $g(y)$  er tæthedsfunktionen for  $Y = t(X)$ . Da udregningens slutresultat ikke involverer  $h$ , vil vi gå ud fra, at det også er lig med selve den søgte sandsynlighed  $P(X \text{ ligger i interval nr. 1} | t(X) = y)$ .

Når betingningsprocessen fordeler sandsynlighedsmassen mellem  $x_1, x_2$  og  $x_3$ , tager den således ikke blot hensyn til tæthedsfunktionen  $f$ s værdier i de tre punkter, men også til, hvor hurtigt  $t$  ændrer sig i de tre punkter.

Som det ses af "eksemplet" kan man sagtens komme ud for, at den betingede fordeling af  $X$  er diskret, selv om  $X$  har en kontinuert fordeling. Mere indviklede situationer opstår når  $X$  er flerdimensional, fordi så vil de betingede fordelinger man kan finde på at interessere sig for typisk være kontinuerte fordelinger. Her indskrænker vi diskussionen til et meget simpelt eksempel.

### Eksempel 2.8. Fordelingen af to eksponentialfordelte variable givet deres sum

Vi fandt tidligere (i Eksempel 2.6) fordelingen af en sum af to uafhængige identisk eksponentialfordelte stokastiske variable  $X_1$  og  $X_2$ . Nu vil vi prøve at bestemme den betingede fordeling af  $X_1$  og  $X_2$  givet deres sum, dvs. givet at  $X_1 + X_2 = y$ . Vi kan først notere, at denne betingede fordeling må være koncentreret på liniestykket  $x_1 + x_2 = y$ ,  $x_1 > 0$ ,  $x_2 > 0$ . Vi kan derfor nøjes med at se på f.eks.  $X_1$ , da  $X_2 = y - X_1$ . For at bestemme den betingede fordeling af  $X_1$  givet at  $X_1 + X_2 = y$  kan man nu først bestemme den betingede sandsynlighed for at  $X_1$  ligger i et  $h$ -interval omkring  $x_1$  givet at  $X_1 + X_2$  ligger i et  $k$ -interval omkring  $y$ , hvor det så er meningen at  $k$  skal gå mod 0. Situationen er illustreret på Figur 2.23. Hvis både  $h$  og  $k$  er små, er sandsynligheden for, at begge betingelserne " $X_1$  ligger i  $h$ -intervallet om  $x_1$ " og " $X_1 + X_2$  ligger i  $k$ -intervallet om  $y$ " er opfyldt, ca. lig med den



Figur 2.23: – vedrørende Eksempel 2.8.

simultane tæthedsfunksions værdi i  $(x_1, x_2) = (x_1, y - x_1)$  ganget med arealet af det skraverede område på figuren<sup>3</sup>; arealet er  $hk$ , så sandsynligheden er dermed ca.

$$\lambda^2 \exp(-\lambda(x_1 + y - x_1)) \times hk = \lambda^2 \exp(-\lambda y) \times hk.$$

Videre er sandsynligheden for at  $Y = X_1 + X_2$  ligger i  $k$ -intervallet omkring  $y$  ca.  $k$  ganget med tæthedsfunktionen for  $Y$ , dvs.  $k\lambda^2 y \exp(-\lambda y)$  ifølge (2.11). Den betingede sandsynlighed for at  $X_1$  tilhører  $h$ -intervallet om  $x_1$  givet at  $X_1 + X_2$  tilhører  $k$ -intervallet om  $y$  er dermed ca.

$$\frac{\lambda^2 \exp(-\lambda y) hk}{\lambda^2 y \exp(-\lambda y) k} = \frac{h}{y}.$$

Når vi lader  $k \searrow 0$  får vi, at den betingede sandsynlighed for at  $X_1$  ligger i  $h$ -intervallet givet at  $X_1 + X_2 = y$  er ca.  $h/y$ , og det betyder at tæthedsfunktionen for  $X_1$  i den betingede fordeling er

$$x_1 \mapsto \begin{cases} 1/y & \text{når } 0 < x_1 < y, \\ 0 & \text{ellers.} \end{cases}$$

Med andre ord, givet at  $X_1 + X_2 = y$  er  $X_1$  ligefordelt på intervallet fra 0 til  $y$ .  $\square$

<sup>3</sup>Det skraverede område er  $\{(x_1, x_2) : x_1 \text{ ligger i } h\text{-intervallet og } x_1 + x_2 \text{ ligger i } k\text{-intervallet}\}$

## Kapitel 3

# Middelværdi, varians, standardafvigelse

Man kan bestemme forskellige karakteristika for stokastiske variable/sandsynlighedsfordelinger. Vi begynder med middelværdi; først kommer en definition, dernæst en omtale af hvad meningen er.

### 3.1 Middelværdi

Situationen er den, at  $X$  er en stokastisk variabel og  $t$  en afbildning der afbilder  $x$ -værdier over i de reelle tal.

**Definition:** Hvis  $X$  er en diskret stokastisk variabel med sandsynlighedsfunktion  $f(x)$ , så defineres *middelværdien* af den stokastiske variabel  $Y = t(X)$  som tallet

$$EY = \sum_x t(x)f(x) .$$

Hvis  $X$  er en kontinuert stokastisk variabel med tæthedsfunktion  $f(x)$ , så defineres *middelværdien* af den stokastiske variabel  $Y = t(X)$  som tallet

$$EY = \int_{-\infty}^{+\infty} t(x)f(x) dx .$$

Hvis specielt afbildningen  $t$  er den identiske afbildning (dvs. afbildningen  $t : x \mapsto x$ ), fås at

*middelværdien* af  $X$  er tallet

$$E X = \sum_x x f(x) ,$$

henholdsvis

$$E X = \int_{-\infty}^{+\infty} x f(x) dx ,$$

hvor  $f$  er sandsynlighedsfunktionen henholdsvis tæthedsfunktionen for  $X$ .

Et andet navn for middelværdi er *forventet værdi*<sup>1</sup>.

Vi vil nu søge at indkredse hvad meningen med middelværdi kan være. Lad os i første omgang tænke på et tilfældigheds"eksperiment" med økonomiske konsekvenser. For nemheds skyld kalder vi det et *lotteri*, men det kan lige så godt handle om, hvorvidt man skal bygge en bro over Store Bælt, eller om man skal tage et års orlov fra studierne eller ej, eller om man skal sammensætte sin aktieportefølje på den ene eller anden eller tredje måde. Lad os sige at lotteriets udfald  $X$  kan antage værdierne  $x_1, x_2, \dots, x_k$  og at de tilsvarende gevinster er  $y_1, y_2, \dots, y_k$ . Hvis man spiller i lotteriet et stort antal  $n$  gange, hvad er så den gennemsnitlige gevinst?

Ifølge hyppighedsfortolkningen af sandsynligheder (side 2) vil udfaldet  $x_j$  optræde i ca. brøkdelen  $P(X = x_j)$  af gangene, så i  $n$  spil vil udfaldet  $x_j$  optræde ca.  $n P(X = x_j)$  gange, og gevinsten i  $n$  spil vil dermed være ca.

$$y_1 n P(X = x_1) + y_2 n P(X = x_2) + \dots + y_k n P(X = x_k)$$

eller

$$\sum_{j=1}^k y_j P(X = x_j) ,$$

---

<sup>1</sup>På engelsk hedder middelværdi (i den her definerede forstand) ligefrem *expected value* eller *expectation*, og det er derfra  $E$ -et stammer.

dvs. netop  $n EY$ . Tallet  $EY$  kan således forstås som den gennemsnitlige gevinst pr. spil i et stort antal spil<sup>2</sup>.

En anden måde at forstå middelværdi på er denne: Antag at man foretager "uafhængige gentagelser" af et bestemt forsøg, dvs. at man observerer uafhængige identisk fordelte stokastiske variable  $Y_1, Y_2, \dots, Y_n$ ; da vil gennemsnittet  $\bar{Y} = (Y_1 + Y_2 + \dots + Y_n)/n$  "med sandsynlighed 1" konvergere mod  $EY$  når  $n \rightarrow \infty$  (iflg. *Store Tals Stærke Lov*). Man kan altså tænke på  $EY$  som det tal hvortil gennemsnit af uafhængige observationer af  $Y$  før eller senere vil nærme sig. Men der foreligger (indtil videre) ikke noget om hvor hurtigt dette sker.

Man kan endelig også finde middelværdien i en fordeling ved at skære det hele ud i pap, sådan rent bogstaveligt. Hvis man tænker sig sandsynlighedsmassen erstattet af sædvanlig tung masse, så svarer middelværdien nemlig til tyngdepunktet i massefordelingen. Man kan derfor finde f.eks. middelværdien i en kontinuert fordeling med tæthed  $f$  ved at tegne grafen for  $f$  på et stykke pap, klippe den ud, og bestemme tyngdepunktets projektion på abscisseaksen, se Figur 3.1.

For en ordens skyld skal vi bemærke, at det faktisk ikke er alle stokastiske variable/sandsynlighedsfordelinger som har en middelværdi; der findes nemlig tilfælde hvor det ikke er muligt at tillægge det matematiske udtryk der skulle bestemme middelværdien en bestemt værdi. Dette diskuteres i det følgende ret lange eksempel; den mindre interesserede læser kan roligt i første omgang springe eksemplet over og fortsætte side 61.

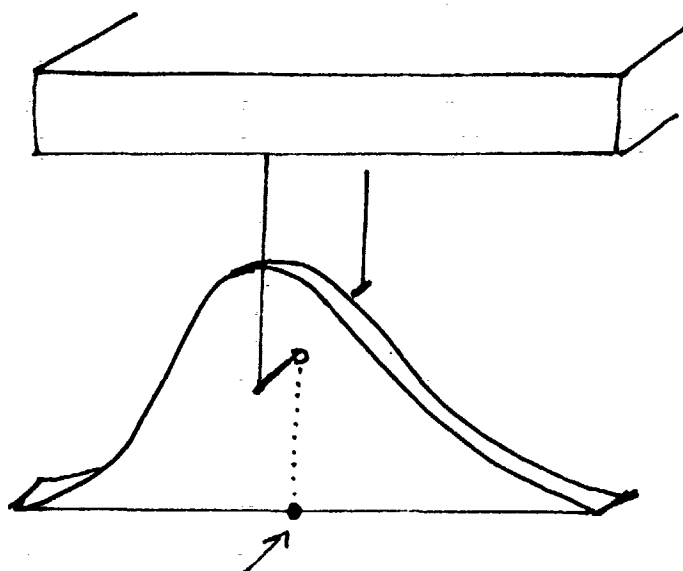
### Eksempel 3.1. En fordeling uden middelværdi

*Antag at vi har en radio-aktiv kilde som udsender  $\alpha$ -partikler på en sådan måde at strålingsintensiteten er den samme i alle retninger. Vi anbringer i et vist tidsrum en (meget stor) fotografisk*

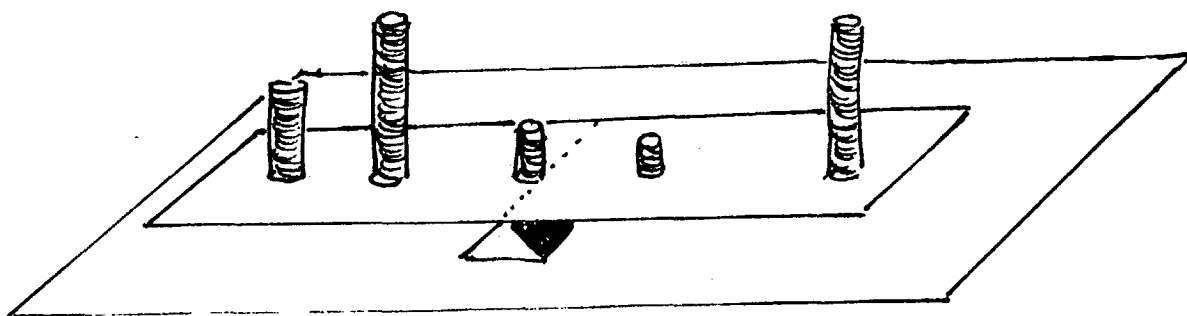
---

<sup>2</sup>Selv om man altså måske på denne måde kan udregne middelværdien af det økonomiske udbytte (det forventede udbytte) af en storebæltsbro, så er det fejlagtigt at tro, at denne middelværdi uden videre fornuftigt kan indgå i vurderingen af, om broprojektet skal iværksættes eller ej. Da man kun vil opleve et enkelt udfald (eller måske to), kan man ikke her udelukkende forlade sig på hvad der vil ske i gennemsnit ved et stort antal broer — under identiske omstændigheder!

Derimod kan eksempelvis et bilforsikringsselskab antagelig godt bygge på sådanne gennemsnitsbetragtninger når det skal fastsætte præmier for de almindelige bilmærker.



Figur 3.1: Eksperimentel bestemmelse af middelværdi.



Figur 3.2: Her er en diskret massefordeling anbragt på et "masseløst" underlag. Balancepunktet er middelværdien i fordelingen.

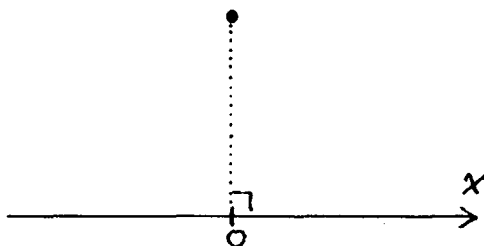
plade et stykke fra kilden for at registrere strålingsmængden i forskellige afstande fra den. Nu er det jo klart, at det område på pladen som er nærmest kilden må modtage flest  $\alpha$ -partikler, men kan man regne ud hvordan strålingsintensiteten vil være på de forskellige dele af pladen?

Vi foretager visse idealiseringer. Vi leger at det hele foregår i to dimensioner, således at den fotografiske "plade" er en ret linie, som i øvrigt er uendelig lang; den radio-aktive kilde er "punktformig"; det hele foregår i lufttomt rum.

• ← 'kilde'

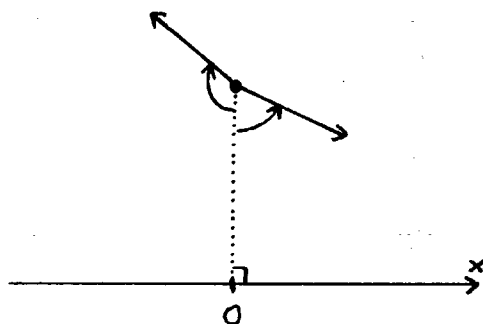
————— ← 'plade'

Vi placerer en tallinie på "pladen" som vist her:

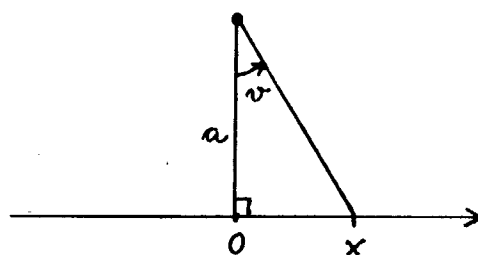


Vi kan beskrive en partikels retning ved at angive den vinkel  $v$  som retningen danner med linien fra "kilden" til talliniens nulpunkt (nulpunktet er valgt så denne linie står vinkelret på tallinien). Vinklen  $v$  opgives som et tal mellem  $-\pi$  og  $\pi$ , således at retninger mod venstre regnes negative. På næste figur er vist to retninger, en med et  $v$  på omkring  $-\frac{3}{4}\pi$  og som ikke rammer "pladen", og

en med et  $v$  på omkring  $\frac{1}{3}\pi$  og som rammer "pladen".



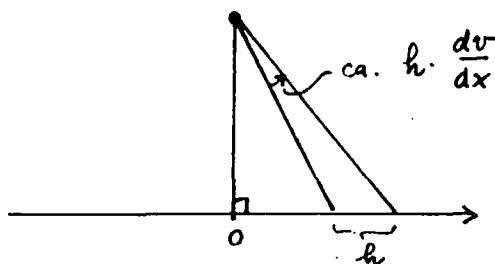
Vi er kun interesserede i de partikler som faktisk rammer, dvs. partikler med et  $v$  mellem  $-\frac{1}{2}\pi$  og  $\frac{1}{2}\pi$ . Da strålingsintensiteten formodes at være den samme i alle retninger, vil vi antage, at vinklen hørende til en partikel som rammer pladen kan antages at være en observation af en stokastisk variabel  $V$  som er ligefordelt på intervallet  $]-\frac{1}{2}\pi, \frac{1}{2}\pi[$ . Imidlertid observerer vi ikke  $V$  men det punkt  $X = X(V)$  hvor partiklen rammer  $x$ -aksen, så vi må bestemme den tilsvarende fordeling af  $X$ .



Sammenhængen mellem  $v$  og  $x$  er den at  $x = a \tan v$ , hvor  $a$  er afstanden mellem kilden og pladen. For at bestemme tæthedsfunktionen for  $X$  finder vi først sandsynligheden for at  $X$  ligger i



et  $h$ -interval omkring  $x$ .



Når  $h$  er lille, er det tilsvarende interval for  $v$  af længde ca.  $h \times \frac{dv}{dx}$ , så tætheden for  $X$  er

$$\begin{aligned} f(x) &= \frac{dv}{dx} \times \text{tætheden for } V \\ &= \frac{1}{\pi} \times \frac{dv}{dx}. \end{aligned}$$

(Dette er blot endnu et eksempel på transformation af tætheder, jf. (2.9) side 41.)

Da  $x = a \tan v$ , er

$$\begin{aligned} \frac{dx}{dv} &= a (1 + (\tan v)^2) \\ &= a (1 + (x/a)^2), \end{aligned}$$

så

$$\frac{dv}{dx} = \frac{1}{a} \times (1 + (x/a)^2)^{-1}.$$

Tætheden for  $X$  er dermed

$$f(x) = \frac{1}{a\pi} \times \frac{1}{1 + (x/a)^2}, \quad -\infty < x < +\infty.$$

Sandsynlighedsfordelingen bestemt ved denne tæthed hedder Cauchy-fordelingen med skalaparameter  $a$ .

Omsider er vi nået til det egentlige: Cauchy-fordelingen er et eksempel på en fordeling som ikke har nogen middelværdi. Det

skal nu vises. Lad os for nemheds skyld sige at  $a = 1$ . Vi skal da se på, om man kan tillægge udtrykket

$$\int_{-\infty}^{+\infty} x f(x) dx = \int_{-\infty}^{+\infty} x \times \frac{1}{\pi} \frac{1}{1+x^2} dx$$

en værdi. Problemerne hænger sammen med spørgsmålet om, hvordan man egentlig skal fortolke et integral hvor grænserne er  $-\infty$  og  $+\infty$ .

**Regel:**  $\int_{-\infty}^{+\infty} g(x) dx$  skal betyde grænseværdien af  $\int_{-A}^B g(x) dx$  når  $A$  og  $B$  på vilkårlig måde går mod  $+\infty$ ; hvis der ikke findes en sådan grænseværdi, så har  $\int_{-\infty}^{+\infty} g(x) dx$  ingen mening.

**Påstand:**

$$\int_0^B \frac{x}{1+x^2} dx \rightarrow +\infty \text{ når } B \rightarrow +\infty ,$$

og

$$\int_{-A}^0 \frac{x}{1+x^2} dx \rightarrow -\infty \text{ når } A \rightarrow +\infty .$$

Vi viser påstanden om lidt; dens konsekvens er, at

$$\int_{-A}^B \frac{x}{1+x^2} dx = \int_{-A}^0 \frac{x}{1+x^2} dx + \int_0^B \frac{x}{1+x^2} dx$$

ikke har nogen grænseværdi når  $A$  og  $B$  på vilkårlig måde går mod  $+\infty$ , hvilket hænger sammen med, at man ikke kan tillægge  $+\infty + (-\infty)$  nogen værdi. Ifølge ovennævnte regel har Cauchyfordelingen derfor ingen middelværdi.

Afslutningsvis må vi vise påstanden; vi nøjes med den første del. Integranden er

$$\frac{x}{1+x^2} = \frac{1}{\frac{1}{x} + x} ,$$

og når  $x$  er stor er det næsten det samme som  $\frac{1}{1+x}$ . Nu ved man at

$$\begin{aligned} \int_0^B \frac{1}{1+x} dx &= \int_1^{1+B} \frac{1}{x} dx \\ &= \ln(1+B) \end{aligned}$$

går mod  $+\infty$  når  $B \rightarrow +\infty$ . Derfor vil

$$\int_0^B \frac{x}{1+x^2} dx \rightarrow +\infty$$

når  $B \rightarrow +\infty$ . Mere præcist kan man sige, at

- når  $x \geq 1$  så er  $\frac{1}{x} \leq 1$  og derfor  $1/(\frac{1}{x} + x) \geq 1/(1+x)$ , og
- når  $0 < x < 1$  er  $1/(\frac{1}{x} + x) \geq 0$ ,

så derfor er

$$\begin{aligned} \int_0^B \frac{1}{\frac{1}{x} + x} dx &\geq \int_1^B \frac{1}{1+x} dx \\ &= \ln(1+B) - \ln 2; \end{aligned}$$

da  $\ln(1+B) - \ln 2 \rightarrow +\infty$  når  $B \rightarrow +\infty$ , vil dermed også

$$\int_0^B \frac{x}{1+x^2} dx \rightarrow +\infty$$

når  $B \rightarrow +\infty$ . □

## 3.2 Regneregler for middelværdier

I det følgende går vi ud fra, at de optrædende stokastiske variable har en middelværdi og at den er et endeligt tal (dvs. ikke  $+\infty$  eller  $-\infty$ ).  
— Der er forskellige nyttige regneregler for middelværdier:

**M1.** Hvis  $X \geq 0$  så er  $EX \geq 0$ .

**M2.**  $Ea = a$ .

**M3.**  $E(X_1 + X_2) = EX_1 + EX_2$ .

**M4.**  $E(aX) = aEX$ .

**M5.** hvis  $X_1$  og  $X_2$  er stokastisk uafhængige, så er  $E(X_1 X_2) = EX_1 EX_2$ .

— Her betegner  $a$  en konstant og  $X, X_1$  og  $X_2$  stokastiske variable.

De fire første egenskaber udtrykkes undertiden på den måde, at middelværdioperationen er en *positiv* (M1), *normeret* (M2), og *lineær* (M3+M4) *funktional* på mængden af stokastiske variable med endelig middelværdi.

Her er skitser til bevis for M1-M5 i fald der er tale om diskrete stokastiske variable; bevis for kontinuerte variable forløber tilsvarende.

$$\text{ad M1: } EX = \sum_x x f(x) = \sum_{x \geq 0} x f(x) \geq 0.$$

$$\text{ad M2: } EX = \sum_x x f(x) = a \times 1 = a.$$

ad M3:

$$\begin{aligned} E(X_1 + X_2) &= \sum_{x_1} \sum_{x_2} (x_1 + x_2) f(x_1, x_2) \\ &= \sum_{x_1} \sum_{x_2} x_1 f(x_1, x_2) + \sum_{x_2} \sum_{x_1} x_2 f(x_1, x_2) \\ &= \sum_{x_1} \left( x_1 \sum_{x_2} f(x_1, x_2) \right) + \sum_{x_2} \left( x_2 \sum_{x_1} f(x_1, x_2) \right) \\ &= \sum_{x_1} x_1 f_1(x_1) + \sum_{x_2} x_2 f_2(x_2) \\ &= EX_1 + EX_2. \end{aligned}$$

$$\text{ad M4: } E(aX) = \sum_x ax f(x) = a \sum_x x f(x) = a EX.$$

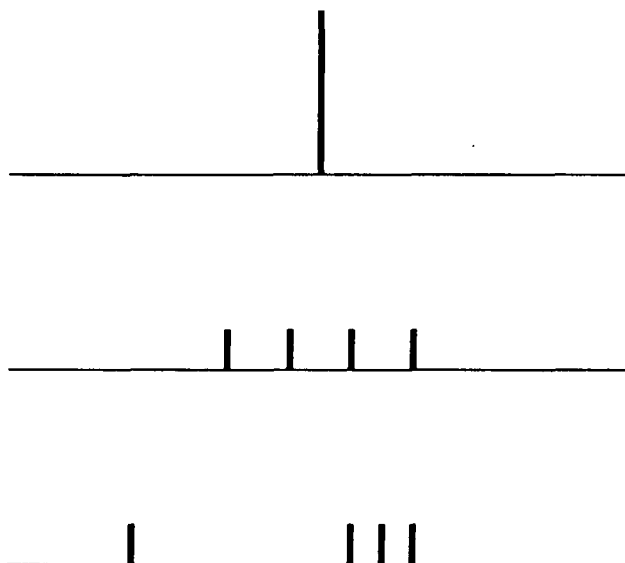
$$\begin{aligned} \text{ad M5: } E(X_1 X_2) &= \sum_{x_1} \sum_{x_2} x_1 x_2 f(x_1, x_2) \\ &= \sum_{x_1} \left( x_1 f_1(x_1) \sum_{x_2} x_2 f_2(x_2) \right) \\ &= \left( \sum_{x_1} x_1 f_1(x_1) \right) \times \left( \sum_{x_2} x_2 f_2(x_2) \right) \\ &= EX_1 EX_2. \end{aligned}$$

### 3.3 Varians

Man kan sige at middelværdien  $EX$  af en stokastisk variabel  $X$  er et forsøg på at approksimere den *stokastiske* variabel  $X$  (eller dens sandsynlighedsfordeling) med et enkelt, *ikke-stokastisk* tal. Middelværdien fortæller, hvor på tallinien fordelingen er beliggende, således at forstå at hvis man forskyder fordelingen stykket  $a$ , dvs. erstatter  $X$  med  $a + X$ , så lægges der  $a$  til middelværdien. Der gælder jo (ifølge M3 og

M2) at  $E(a + X) = a + E X$ , hvilket med et fint ord udtrykkes som at middelværldioperationen er *translations-ækvivariant*.

Men middelværdien er langt fra altid tilstrækkelig til at beskrive fordelingen; her er vist et par diskrete fordelinger som alle har samme middelværdi:



Man kan så sætte sig som opgave at beskrive fordelingen ved hjælp af *to* tal, hvoraf det ene skal være middelværdien og det andet skal beskrive fordelingsens spredning omkring middelværdien. Som det andet tal plejer man at bruge enten variansen eller standardafvigelsen (som er kvadratroden af variansen). Variansen på  $X$  er den forventede kvadratiske afvigelse mellem  $X$  og  $E X$ :

**Variansen** af en stokastisk variabel  $X$  er tallet

$$\text{Var } X = E(X - E X)^2.$$

**Standardafvigelsen** af  $X$  er tallet  $\sqrt{\text{Var } X}$ .

Her skal udtrykket  $E(X - E X)^2$  forstås som  $E t(X)$  hvor  $t$  er funktionen  $t : x \mapsto (x - E X)^2$ .

Det bemærkes at standardafvigelsen på  $X$  måles i samme enheder som  $X$  (f.eks. m eller kg etc.).

Her er forskellige egenskaber og regneregler for varianser:

**V1.**  $\text{Var } X \geq 0$ .

**V2.**  $\text{Var } a = 0$ .

**V3.**  $\text{Var}(aX) = a^2 \text{Var } X$ .

**V4.**  $\text{Var}(a + X) = \text{Var } X$ .

– Her betegner  $a$  en konstant og  $X$  en stokastisk variabel.

V4 udtrykker at variansen (og dermed også standardafvigelsen) er *translations-invariant*, dvs. den afhænger ikke af, hvor på tallinien fordelingen er beliggende. V3 viser at standardafvigelsen på  $aX$  er lig  $a$  gange standardafvigelsen på  $X$  (hvis  $a \geq 0$ ), dvs. standardafvigelsen er *skala-ækvivariant*.

Her er bevisskitser til V1-V4:

**ad V1:** benyt M1 på den positive stokastiske variabel  $(X - EX)^2$ .

**ad V2:**  $E(a - Ea)^2 = E 0 = 0$  ifølge M2.

**ad V3:**  $E((aX - E(aX))^2) = E(aX - aEX)^2 = a^2 E(X - EX)^2$ .

**ad V4:**  $E(a + X - E(a + X))^2 = E(X - EX)^2$ .

Undertiden har man glæde af nogle alternative udtryk for  $\text{Var } X$ :

**V5.**  $\text{Var } X = EX^2 - (EX)^2$ .

**V6.**  $\text{Var } X = EX(X - 1) - (EX)(EX - 1)$ .

$$\begin{aligned}
 (\text{Bevis:} \quad \text{Var } X &= E(X - EX)^2 \\
 &= E(X^2 - 2XEX + (EX)^2) \\
 &= EX^2 - 2(EX)^2 + (EX)^2 \\
 &= EX^2 - (EX)^2,
 \end{aligned}$$

og

$$\begin{aligned}
 EX^2 - (EX)^2 &= (EX^2 - EX) - ((EX)^2 - EX) \\
 &= E(X^2 - X) - (EX)(EX - 1) \\
 &= EX(X - 1) - (EX)(EX - 1) . .
 \end{aligned}$$

Der gælder at middelværdien af en sum er lig summen af middelværdierne (M3). Noget tilsvarende gælder *ikke* for varianser! Man kan godt gå i gang med at omskrive  $\text{Var}(X_1 + X_2)$ , men det bliver ikke noget "pænt":

$$\begin{aligned}\text{Var}(X_1 + X_2) &= E\left((X_1 + X_2) - E(X_1 + X_2)\right)^2 \\ &= E\left((X_1 - E X_1) + (X_2 - E X_2)\right)^2 \\ &= E\left((X_1 - E X_1)^2 + (X_2 - E X_2)^2\right. \\ &\quad \left.+ 2(X_1 - E X_1)(X_2 - E X_2)\right) \\ &= \text{Var} X_1 + \text{Var} X_2 + 2 E(X_1 - E X_1)(X_2 - E X_2) .\end{aligned}$$

Variansen på summen er altså summen af varianserne plus noget mere. Dette "mere" har fået sit eget navn, nemlig *kovariansen* mellem  $X_1$  og  $X_2$ . I denne forbindelse kan vi også nævne at man definerer *korrelationen* mellem  $X_1$  og  $X_2$  som kovariansen divideret med produktet af standardafvigelse:ne:

**Kovariansen** mellem  $X_1$  og  $X_2$  er tallet

$$\text{Cov}(X_1, X_2) = E(X_1 - E X_1)(X_2 - E X_2) .$$

**Korrelationen** mellem  $X_1$  og  $X_2$  er tallet

$$\text{corr}(X_1, X_2) = \frac{\text{Cov}(X_1, X_2)}{\sqrt{\text{Var} X_1} \sqrt{\text{Var} X_2}} .$$

Efter at have indført kovarianser kan man uden videre opskrive nedenstående regel V7 for variansen på en sum, og idet vi forudskikker at uafhængige stokastiske variable har kovarians 0 (regel C7) får vi også regel V8:

**V7.**  $\text{Var}(X_1 + X_2) = \text{Var} X_1 + \text{Var} X_2 + 2 \text{Cov}(X_1, X_2)$ .

**V8.** Hvis  $X_1$  og  $X_2$  er stokastisk uafhængige, så er  $\text{Var}(X_1 + X_2) = \text{Var} X_1 + \text{Var} X_2$ .

Hvor middelværdi, varians og standardafvigelse er karakteristika for en *enkelt* stokastisk variabel/sandsynlighedsfordeling på den reelle akse, er kovarians og korrelation størrelser der beskriver træk ved den *simultane* fordeling af to stokastiske variable (eller én todimensional variabel). Det kan i øvrigt noteres, at kovariansen er et benævnt tal (ligesom middelværdi, varians og standardafvigelse): hvis f.eks.  $X_1$  angives i kg og  $X_2$  i m, så har  $\text{Cov}(X_1, X_2)$  benævnelsen kgm. Derimod er korrelationen et ubenævnt tal, som i øvrigt altid vil være mellem  $-1$  og  $1$ .

Om kovarianser gælder at

- en *positiv* kovarians er tegn på at store  $X_1$ -værdier (i.e.  $X_1 > E X_1$ ) fortrinsvis optræder sammen med store  $X_2$ -værdier (i.e.  $X_2 > E X_2$ ), og små  $X_1$ -værdier ( $X_1 < E X_1$ ) fortrinsvis sammen med små  $X_2$ -værdier ( $X_2 < E X_2$ ).
- en *negativ* kovarians er tegn på at store  $X_1$ -værdier ( $X_1 > E X_1$ ) fortrinsvis optræder sammen med små  $X_2$ -værdier ( $X_2 < E X_2$ ), og små  $X_1$ -værdier ( $X_1 < E X_1$ ) fortrinsvis sammen med store  $X_2$ -værdier ( $X_2 > E X_2$ ).

De følgende egenskaber ved og regneregler for kovarianser vises (med undtagelse af C7) let, som oftest på lignende måde som de tilsvarende regler V1-V6; C7 følger af C6 og M5.

C1.  $\text{Cov}(X, X) = \text{Var} X.$

C2.  $\text{Cov}(X_1, X_2) = \text{Cov}(X_2, X_1).$

C3.  $\text{Cov}(a, X) = 0.$

C4.  $\text{Cov}(a + X_1, b + X_2) = \text{Cov}(X_1, X_2).$

C5.  $\text{Cov}(aX_1, bX_2) = ab \text{Cov}(X_1, X_2).$

C6.  $\text{Cov}(X_1, X_2) = E(X_1 X_2) - E X_1 E X_2.$

C7. Hvis  $X_1$  og  $X_2$  er stokastisk uafhængige, så er  $\text{Cov}(X_1, X_2) = 0.$

– Her betegner  $a$  og  $b$  konstanter og  $X, X_1$  og  $X_2$  stokastiske variable.



## Kapitel 4

### Nogle resultater om normalfordelingen

I statistik benytter man sig tit af normalfordelingsmodeller. I dette kapitel skal vi indføre normalfordelingen og nogle deraf afledte fordelinger ( $\chi^2$ -fordelingen,  $t$ -fordelingen,  $F$ -fordelingen). Ved at anvende teknikker og resultater fra de tidligere kapitler kan man udlede nogle vigtige resultater om normalfordelingen og de deraf afledte fordelinger. Som en lidt større matematisk øvelse udledes endvidere tætheden for  $\chi^2$ -fordelingen.

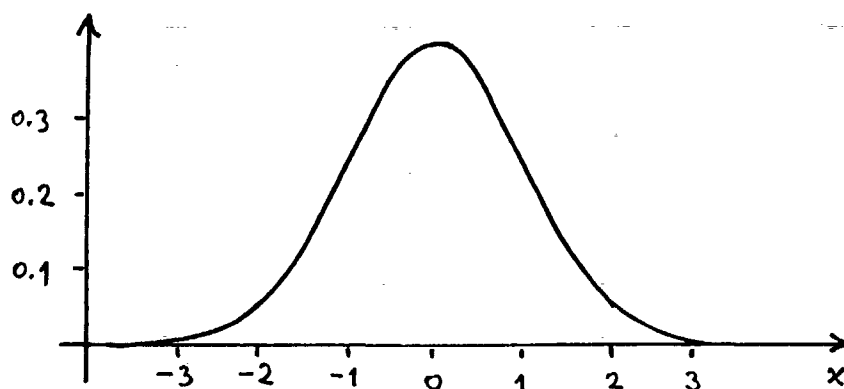
#### 4.1 Normalfordelingen

Vi vil i første omgang definere *normalfordelingen med positionsparameter  $\mu$  og kvadratisk skalaparameter  $\sigma^2$* , kort  $\mathcal{N}(\mu, \sigma^2)$ , som den sandsynlighedsfordeling på den reelle akse hvis tæthedsfunktion er proportional med

$$\exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right).$$

Proportionalitetsfaktoren skal bestemmes således at tæthedsfunktionen integrerer til 1, så vi kan skrive tæthedsfunktionen  $f(x; \mu, \sigma^2)$  for normalfordelingen  $\mathcal{N}(\mu, \sigma^2)$  som

$$f(x; \mu, \sigma^2) = \frac{1}{c(\mu, \sigma^2)} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right), \quad -\infty < x < +\infty,$$



Figur 4.1: Tætheden for  $\mathcal{N}(0, 1)$ .

hvor

$$c(\mu, \sigma^2) = \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2} \frac{(x - \mu)^2}{\sigma^2}\right) dx. \quad (4.1)$$

For at gøre definitionen lidt pænere må vi finde ud af, hvad  $c(\mu, \sigma^2)$  nærmere er. Det gør vi på de følgende sider, samtidig med at vi udleder forskellige vigtige resultater om normalfordelingen.

Først vises at størrelsen  $c(\mu, \sigma^2)$  slet ikke afhænger af  $\mu$  og at den afhænger af  $\sigma^2$  på en simpel måde.

Hvis man i integraludtrykket (4.1) for  $c(\mu, \sigma^2)$  foretager substitutionen  $z = (x - \mu)/\sigma$ , så får man at

$$\begin{aligned} c(\mu, \sigma^2) &= \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2} \frac{(x - \mu)^2}{\sigma^2}\right) dx \\ &= \sigma \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2} z^2\right) dz, \end{aligned}$$

så størrelsen  $c(\mu, \sigma^2)$  er "bare"  $c\sigma$ , hvor  $c$  er tallet  $\int_{-\infty}^{+\infty} \exp(-\frac{1}{2} z^2) dz$ . Tæthedsfunktionen for  $\mathcal{N}(\mu, \sigma^2)$  er således "bare"

$$f(x; \mu, \sigma^2) = \frac{1}{c\sigma} \exp\left(-\frac{1}{2} \frac{(x - \mu)^2}{\sigma^2}\right).$$

Inden vi bestemmer tallet  $c$ , vil vi vise hvordan normalfordelingen "transformerer ved affine transformationer", dvs. vi vil vise hvad der

sker, når man transformerer en  $\mathcal{N}(\mu, \sigma^2)$ -stokastisk variabel  $X$  med den affine transformation  $y = ax + b$ . Ifølge den generelle formel for transformation af sandsynlighedstætheder (side 41) er (når  $a \neq 0$ ) tætheden for  $Y = aX + b$  givet ved

$$\begin{aligned} g(y) &= f\left(\frac{y-b}{a}; \mu, \sigma^2\right) \left|\frac{1}{a}\right| \\ &= \frac{1}{|a|c\sigma} \exp\left(-\frac{1}{2} \frac{(y-b-a\mu)^2}{a^2\sigma^2}\right) \\ &= f(y; a\mu + b, a^2\sigma^2), \end{aligned}$$

så dermed har vi eftervist at<sup>a</sup>

- Hvis  $X$  er  $\mathcal{N}(\mu, \sigma^2)$ -fordelt, så er  $aX + b$   $\mathcal{N}(a\mu + b, a^2\sigma^2)$ -fordelt.

Som et specialtilfælde af denne regel har vi

- Hvis  $X$  er  $\mathcal{N}(0, 1)$ -fordelt, så er  $\sigma X + \mu$   $\mathcal{N}(\mu, \sigma^2)$ -fordelt.
- Hvis  $X$  er  $\mathcal{N}(\mu, \sigma^2)$ -fordelt, så er  $(X - \mu)/\sigma$   $\mathcal{N}(0, 1)$ -fordelt.

Tæthedsfunktionen  $f(\cdot; 0, 1)$  for den "normerede normale fordeling"  $\mathcal{N}(0, 1)$  betegnes ofte  $\phi$ :

$$\phi(x) = \frac{1}{c} \exp\left(-\frac{1}{2}x^2\right).$$

Med denne betegnelse kan tæthedsfunktionen for den generelle normalfordeling skrives som

$$f(x; \mu, \sigma^2) = \frac{1}{\sigma} \phi\left(\frac{x - \mu}{\sigma}\right).$$

---

<sup>a</sup>Det er på grund af denne egenskab at  $\mu$  kaldes en *positionsparameter* og  $\sigma^2$  en *kvadratisk skalaparameter*.

### Bestemmelse af $c$

Vi vil nu bestemme tallet  $c$  ved et snedigt trick.  $c$  er defineret som normeringsfaktoren i  $\mathcal{N}(0, 1)$ -fordelingen, men tricket består i at se på *to* stokastisk uafhængige  $\mathcal{N}(0, 1)$ -variable: i deres simultane todimensionale fordeling er normeringsfaktoren  $c^2$ , og det viser sig at  $c^2$  er let at bestemme. Vi betragter derfor  $X_1$  og  $X_2$  der skal være stokastisk uafhængige  $\mathcal{N}(0, 1)$ -variable. Deres simultane tæthedsfunktion er

$$\begin{aligned} f(x_1, x_2) &= \phi(x_1) \phi(x_2) \\ &= \frac{1}{c^2} \exp\left(-\frac{1}{2}(x_1^2 + x_2^2)\right). \end{aligned}$$

Normeringsfaktoren  $c^2$  skal bevirke at denne todimensionale tæthed integrerer til 1.

Vi ser nu på en transformeret stokastisk variabel  $(Y_1, Y_2) = t(X_1, X_2)$ , hvor  $Y_1 = X_1^2 + X_2^2$  og hvor  $Y_2$  er vinklen fra abscisseaksen til linien fra  $(0, 0)$  gennem  $(X_1, X_2)$ . Det er den transformation som studeredes indgående i Eksempel 2.7, hvor vi fandt at

$$\begin{aligned} X_1 &= Y_1^{1/2} \cos Y_2 \\ X_2 &= Y_1^{1/2} \sin Y_2, \end{aligned}$$

og at den simultane tæthedsfunktion  $g(y_1, y_2)$  for  $Y_1$  og  $Y_2$  er, når  $y_1 > 0$  og  $0 \leq y_2 < 2\pi$ ,

$$\begin{aligned} g(y_1, y_2) &= \frac{1}{2} f(x_1(y_1, y_2), x_2(y_1, y_2)) \\ &= \frac{1}{2c^2} \exp\left(-\frac{1}{2}(x_1^2 + x_2^2)\right) \\ &= \frac{1}{2c^2} \exp\left(-\frac{1}{2}y_1\right). \end{aligned} \tag{4.2}$$

$g$  er fremkommet på en måde så vi véd at den er en sandsynlighedstæthed, og derfor véd vi også helt automatisk at

$$1 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(y_1, y_2) dy_2 dy_1.$$

Når man indsætter udtrykket for  $g$  og omskriver får man

$$1 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(y_1, y_2) dy_2 dy_1$$

$$\begin{aligned}
&= \int_0^{+\infty} \left( \int_0^{2\pi} \frac{1}{2c^2} \exp(-\tfrac{1}{2}y_1) dy_2 \right) dy_1 \\
&= \frac{1}{c^2} \int_0^{+\infty} \exp(-\tfrac{1}{2}y_1) \left( \int_0^{2\pi} 1 dy_2 \right) dy_1 \\
&= \frac{2\pi}{c^2} \int_0^{+\infty} \exp(-\tfrac{1}{2}y_1) dy_1 \\
&= \frac{2\pi}{c^2}.
\end{aligned}$$

Da således  $1 = 2\pi/c^2$  må  $c$  være lig med  $\sqrt{2\pi}$ , så omsider kender vi tætheden for  $\mathcal{N}(\mu, \sigma^2)$ :

Tætheden for  $\mathcal{N}(\mu, \sigma^2)$ -fordelingen er

$$\frac{1}{\sqrt{2\pi}\sigma^2} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right), \quad -\infty < x < +\infty.$$

Et andet udtryk for tætheden er

$$\frac{1}{\sigma} \phi\left(\frac{x-\mu}{\sigma}\right)$$

hvor

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp(-\tfrac{1}{2}x^2)$$

er tætheden for den normerede normale fordeling  $\mathcal{N}(0, 1)$ .

## Middelværdi og varians

Vi vil nu bestemme middelværdi og varians af normalfordelte stokastiske variable, i første omgang af en  $\mathcal{N}(0, 1)$ -variabel  $X$ .

Da  $\mathcal{N}(0, 1)$ -fordelingen er symmetrisk omkring 0, må man formode at dens middelværdi er 0. Det er da også tilfældet, men den læser som har studeret Eksempel 3.1 vil vide, at man godt kan blive narret når det handler om integraler hvor grænserne er  $-\infty$  og  $+\infty$ . Vi vil derfor give et rigtigt bevis for, at  $EX = 0$ :

Ifølge definitionen er

$$E X = \int_{-\infty}^{+\infty} x \phi(x) dx .$$

Med skæringspunkt  $x = 0$  deles integrationsområdet op i to dele som behandles hver for sig. Ved at foretage substitutionen  $z = x^2$  fås

$$\begin{aligned} \int_0^{+\infty} x \phi(x) dx &= \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} z^{1/2} \exp(-\tfrac{1}{2}z) \tfrac{1}{2} z^{-1/2} dz \\ &= \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} \tfrac{1}{2} \exp(-\tfrac{1}{2}z) dz \\ &= \frac{1}{\sqrt{2\pi}} . \end{aligned}$$

Tilsvarende er  $\int_{-\infty}^0 x \phi(x) dx = -1/\sqrt{2\pi}$ , så

$$\begin{aligned} E X &= \frac{1}{\sqrt{2\pi}} + \frac{-1}{\sqrt{2\pi}} \\ &= 0 \end{aligned}$$

som formodet.

Da den  $\mathcal{N}(0, 1)$ -fordelte stokastiske variabel  $X$  har middelværdi 0, gælder om dens varians at

$$\begin{aligned} \text{Var } X &= E X^2 \\ &= \int_{-\infty}^{+\infty} x^2 \phi(x) dx \\ &= 2 \int_0^{\infty} x^2 \phi(x) dx . \end{aligned}$$

Opmuntret af successen fra før prøver vi også her med substitutionen  $z = x^2$ :

$$\begin{aligned} E X^2 &= \frac{2}{\sqrt{2\pi}} \int_0^{+\infty} z \exp(-\tfrac{1}{2}z) \tfrac{1}{2} z^{-1/2} dz \\ &= \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} z^{1/2} \exp(-\tfrac{1}{2}z) dz , \end{aligned}$$

men dette integral lader sig ikke bestemme ved standardmetoder. Vi benytter derfor et nyt snedigt trick, eller rettere: vi går videre med

trick'et fra afsnittet om bestemmelse af  $c$ . Lad  $X_1$  og  $X_2$  være uafhængige  $\mathcal{N}(0, 1)$ -variable, og lad  $Y_1$  og  $Y_2$  være som på side 70. Specielt er  $Y_1 = X_1^2 + X_2^2$  og dermed  $E Y_1 = E X_1^2 + E X_2^2 = 2 E X_1^2$ ; det søgte tal  $\text{Var} X_1 = E X_1^2$  er derfor halvdelen af tallet  $E Y_1$  som vi vil bestemme. Den simultane tæthedsfunktion  $g(y_1, y_2)$  for  $Y_1$  og  $Y_2$  har fandt vi tidligere til at være givet ved (4.2). Den marginale tæthed for  $Y_1$  er da (jf. (2.7), side 36) for  $y_1 > 0$

$$\begin{aligned} g_1(y_1) &= \int_{-\infty}^{+\infty} g(y_1, y_2) dy_2 \\ &= \int_0^{2\pi} \frac{1}{4\pi} \exp(-\tfrac{1}{2}y_1) dy_2 \\ &= \tfrac{1}{2} \exp(-\tfrac{1}{2}y_1), \end{aligned} \tag{4.3}$$

dvs.  $Y_1$  er eksponentialfordelt med parameter  $\frac{1}{2}$ . Middelværdien af  $Y_1$  kan herefter let findes ved partiel integration:

$$\begin{aligned} E Y_1 &= \int_0^{+\infty} y g_1(y) dy \\ &= \int_0^{+\infty} y \tfrac{1}{2} \exp(-\tfrac{1}{2}y) dy \\ &= \left[ -y \exp(-\tfrac{1}{2}y) \right]_0^{+\infty} + \int_0^{+\infty} \exp(-\tfrac{1}{2}y) dy \\ &= 0 + 2 \\ &= 2. \end{aligned}$$

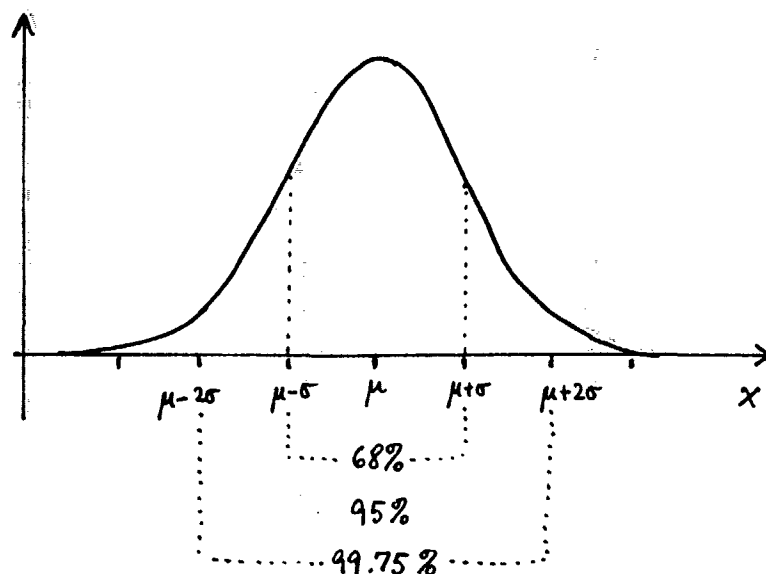
Den søgte varians i  $\mathcal{N}(0, 1)$ -fordelingen er derfor lig 1.

Hvis  $X$  er en  $\mathcal{N}(0, 1)$ -fordelt stokastisk variabel, så er  $E X = 0$  og  $\text{Var} X = 1$ .

Som tidligere vist gælder, at hvis  $X$  er  $\mathcal{N}(0, 1)$ -fordelt, så er  $\mu + \sigma X$   $\mathcal{N}(\mu, \sigma^2)$ -fordelt. Heraf og af regnereglerne for middelværdi og varians fås

Hvis  $X$  er en  $\mathcal{N}(\mu, \sigma^2)$ -fordelt stokastisk variabel, så er  $E X = \mu$  og  $\text{Var} X = \sigma^2$ .

Normalfordelingens standardafvigelse  $\sigma$  er en hensigtsmæssig enhed at benytte når man skal beskrive egenskaber ved normalfordelingskurven. Der gælder at



Figur 4.2:  $\sigma$ s relation til normalfordelingskurven.

- intervallet  $\mu \pm 2\sigma$  indeholder ca. 95% af sandsynlighedsmassen<sup>b</sup>,
- intervallet  $\mu \pm 3\sigma$  indeholder ca. 99.75% af sandsynlighedsmassen, og
- intervallet  $\mu \pm \sigma$  indeholder ca. 68% af sandsynlighedsmassen,

se også Figur 4.2.

## 4.2 $\chi^2$ -fordelingen

Man benytter meget ofte statistiske modeller som går ud på at observationer er normalfordelte. I den forbindelse kommer man ud for at skulle bestemme fordelingerne af visse funktioner af normalfordelte

<sup>b</sup>Mere præcist: 95% af sandsynlighedsmassen ligger i intervallet  $\mu \pm 1.96\sigma$ ; tallet 1.96 fremkommer som 97.5%-fraktilen i  $\mathcal{N}(0, 1)$ -fordelingen.



stokastiske variable. Den såkaldte  $\chi^2$ -fordeling (khi-i-anden fordeling) er således fordelingen af en sum af kvadrerede  $\mathcal{N}(0, 1)$ -variable:

**Definition:** Hvis  $X_1, X_2, \dots, X_k$  er stokastisk uafhængige  $\mathcal{N}(0, 1)$ -variable, så er kvadratsummen  $Y = X_1^2 + X_2^2 + \dots + X_k^2$   $\chi^2$ -fordelt med  $k$  frihedsgrader.

Det fremgår umiddelbart, at

- Hvis  $Y_1$  og  $Y_2$  er stokastisk uafhængige og  $\chi^2$ -fordelte med hhv.  $j$  og  $k$  frihedsgrader, så er  $Y_1 + Y_2$   $\chi^2$ -fordelt med  $j + k$  frihedsgrader,

idet

$$Y_1 + Y_2 = \underbrace{X_1^2 + X_2^2 + \dots + X_j^2}_{Y_1} + \underbrace{X_{j+1}^2 + X_{j+2}^2 + \dots + X_{j+k}^2}_{Y_2}.$$

Det er muligt at bevise at

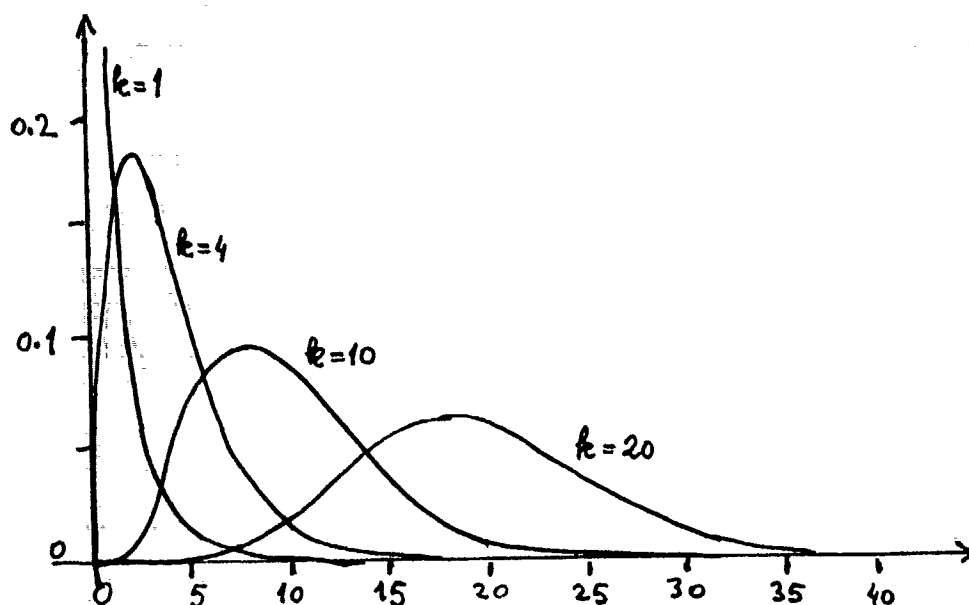
- Sandsynlighedstæthedsfunktionen for  $\chi^2$ -fordelingen med  $k$  frihedsgrader er

$$f_k(y) = \begin{cases} \frac{1}{(\frac{k}{2} - 1)!} \left(\frac{1}{2}\right)^{\frac{k}{2}} y^{\frac{k}{2}-1} \exp(-\frac{1}{2}y), & y > 0 \\ 0 & \text{ellers.} \end{cases}$$

I udtrykket for  $f_k$  optræder der (hvis  $k$  er ulige) fakultetsfunktionen for halvtal, så der er behov for en udvidelse af fakultetsfunktionen

**Definition:** Hvis  $x$  er et heltal eller halvtal større end  $-1$ , så defineres  $x!$  ved

$$\begin{aligned} x! &= x(x-1)! \text{ når } x > 0, \\ 0! &= 1, \\ \left(-\frac{1}{2}\right)! &= \sqrt{\pi}. \end{aligned}$$



Figur 4.3: Tætheder for nogle  $\chi^2$ -fordelinger.

Vi har allerede tidligere uforvarende fundet nogle  $\chi^2$ -tætheder. Et par gange i det foregående har vi set på  $X_1^2 + X_2^2$  hvor  $X_1$  og  $X_2$  er uafhængige  $\mathcal{N}(0, 1)$ -variable. Vi fandt blandt andet (i (4.3)) tætheden for  $Y = X_1^2 + X_2^2$  til  $\frac{1}{2} \exp(-\frac{1}{2}y)$ , hvilket netop er  $f_2(y)$ .  $\chi^2$ -fordelingen med to frihedsgrader er altså det samme som eksponentialfordelingen med parameter  $\lambda = \frac{1}{2}$ .  $\chi^2$ -fordelingen med fire frihedsgrader er derfor det samme som fordelingen af en sum af to uafhængige eksponentialfordelte størrelser hver med parameter  $\frac{1}{2}$ ; vi fandt i Eksempel 2.6 at denne fordeling havde tæthedsfunktionen  $(\frac{1}{2})^2 y \exp(-\frac{1}{2}y)$ , altså netop  $f_4(y)$ .

Tætheden for  $\chi^2$ -fordelingen med én frihedsgrad er let at bestemme ud fra 'first principles'. Det handler om at bestemme fordelingen af  $Y = X^2$  hvor  $X$  er  $\mathcal{N}(0, 1)$ , dvs. det er en simpel transformationsopgave. Man får, jf. side 41, at tætheden for  $Y$  er

$$\begin{aligned} g(y) &= \frac{\phi(x)}{2x} + \frac{\phi(-x)}{2x} \\ &= \frac{\phi(x)}{x}, \end{aligned}$$

hvor  $x = y^{1/2}$  og  $\phi$  som sædvanlig betegner tætheden for  $\mathcal{N}(0, 1)$ . Med andre ord,

$$g(y) = \frac{1}{\sqrt{\pi}} \left(\frac{1}{2}\right)^{1/2} y^{-1/2} \exp\left(-\frac{1}{2}y\right), \quad y > 0,$$

altså  $g(y) = f_1(y)$ .

Vi har således bevist påstanden om  $\chi^2$ -fordelingens tæthed for  $k = 1, 2$  og  $4$ . For fuldstændighedens skyld giver vi nu et ordentligt bevis for den generelle påstand (og undervejs udledes et par integralformler). Den læser der ikke måtte være oplagt til et sådant bevis kan springe det over og læse videre side 79.

Først noterer vi, at hvis  $x$  er et ikke-negativt heltal, så er

$$x! = \int_0^{+\infty} t^x \exp(-t) dt.$$

Thi ved partiel integration får man for  $x > 0$

$$\begin{aligned} \int_0^{+\infty} t^x \exp(-t) dt &= [-t^x \exp(-t)]_0^{+\infty} + \int_0^{+\infty} x t^{x-1} \exp(-t) dt \\ &= 0 + x \int_0^{+\infty} t^{x-1} \exp(-t) dt, \end{aligned}$$

der ved gentagen anvendelse giver

$$\begin{aligned} \int_0^{+\infty} t^x \exp(-t) dt &= x(x-1)(x-2)\dots 2 \times 1 \int_0^{+\infty} t^0 \exp(-t) dt \\ &= x! \end{aligned}$$

idet  $\int_0^{+\infty} \exp(-t) dt = 1$ .

Integraludtrykket for  $x!$  er defineret også selv om  $x$  ikke er et heltal, blot skal  $x$  være større end  $-1$ . Man kan derfor vælge at *definere*  $x!$  for vilkårligt  $x > -1$  ved formlen<sup>c</sup>

$$x! = \int_0^{+\infty} t^x \exp(-t) dt, \quad x > -1. \quad (4.4)$$

<sup>c</sup>For almindannelsens skyld nævner vi, at den såkaldte *gammafunktion* (eller  *$\Gamma$ -funktion*) defineres som

$$\Gamma(x) = \int_0^{+\infty} t^{x-1} \exp(-t) dt, \quad x > 0,$$

dvs.  $\Gamma(x) = (x-1)!$ .

Den just foretagne omskrivning med partiel integration giver da at

$$x! = x(x-1)!, \quad x > 0. \quad (4.5)$$

Vi har brug for at kende  $x!$  for halvtallige  $x$ . Hvis vi kender  $(-\frac{1}{2})!$ , kan vi ved hjælp af (4.4) bestemme værdierne for alle de andre halvtal. Men ved at foretage substitutionen  $t = \frac{1}{2}x^2$  fås

$$\begin{aligned} (-\tfrac{1}{2})! &= \int_0^{+\infty} t^{-1/2} \exp(-\tfrac{1}{2}t) dt \\ &= \sqrt{2} \int_0^{+\infty} \exp(-\tfrac{1}{2}u^2) du \\ &= \sqrt{2} \cdot \tfrac{1}{2} \sqrt{2\pi} \\ &= \sqrt{\pi}. \end{aligned}$$

Hermed har vi godtgjort den udvidede definition (4.4) af  $x!$ .

For hvert heltal  $k$  er funktionen  $f_k$  en sandsynlighedstæthed, thi for det første er øjensynlig  $f_k(y) \geq 0$  for alle  $y$ , og for det andet er  $\int_0^{+\infty} f_k(y) dy = 1$ , hvilket følger umiddelbart ved at foretage substitutionen  $y = 2t$  i integralet.

Vi vil nu vise, at hvis  $Y_1$  og  $Y_2$  er uafhængige stokastiske variable med tæthedsfunktion hhv.  $f_j(y)$  og  $f_k(y)$ , så er tæthedsfunktionen for  $Y_1 + Y_2$  lig  $f_{j+k}(y)$ . Deraf følger nemlig den ønskede påstand om at  $f_k$  er tætheden for  $\chi^2$ -fordelingen med  $k$  frihedsgrader, fordi dels vides påstanden at være rigtig for  $k = 1$ , dels kan  $\chi^2$ -fordelingen med  $k$  frihedsgrader fås som fordelingen af en sum af  $k$  uafhængige  $\chi^2$ -størrelser hver med én frihedsgrad (side 75). Antag altså at  $Y_1$  og  $Y_2$  er stokastisk uafhængige med tætheder hhv.  $f_j(y)$  og  $f_k(y)$ . Tætheden  $g$  for  $Y_1 + Y_2$  er ifølge den generelle formel (2.12) på side 47 for tætheden for en sum givet som

$$g(y) = \int_0^y f_j(z) f_k(y-z) dz.$$

Ved at indsætte udtrykkene for  $f_j$  og  $f_k$  får vi

$$g(y) = \frac{1}{(\frac{j}{2}-1)!(\frac{k}{2}-1)!} \left(\frac{1}{2}\right)^{\frac{j+k}{2}} \exp(-\frac{1}{2}y) \int_0^y z^{\frac{j}{2}-1} (y-z)^{\frac{k}{2}-1} dz.$$

Integralet omskrives på følgende måde:

$$\begin{aligned}\int_0^y z^{\frac{j}{2}-1} (y-z)^{\frac{k}{2}-1} dz &= y^{\frac{j+k}{2}-2} \int_0^y \left(\frac{z}{y}\right)^{\frac{j}{2}-1} \left(1-\frac{z}{y}\right)^{\frac{k}{2}-1} dz \\ &= y^{\frac{j+k}{2}-1} \int_0^1 t^{\frac{j}{2}-1} (1-t)^{\frac{k}{2}-1} dt \\ &= y^{\frac{j+k}{2}-1} \times b(j, k),\end{aligned}$$

hvor  $b(j, k)$  er en betegnelse for det sidste integral. Tætheden for  $Y_1 + Y_2$  synes dermed at være

$$g(y) = \frac{b(j, k)}{(\frac{j}{2}-1)! (\frac{k}{2}-1)!} \left(\frac{1}{2}\right)^{\frac{j+k}{2}} y^{\frac{j+k}{2}-1} \exp\left(-\frac{1}{2}y\right).$$

Vi kan nu se at  $g(y) = \text{konstant} \times f_{j+k}(y)$ , og da både  $g(y)$  og  $f_{j+k}(y)$  er sandsynlighedstætheder, må konstanten være 1. Dermed har vi vist at  $Y_1 + Y_2$  har tæthed  $f_{j+k}(y)$ .

Da  $g(y) = f_{j+k}(y)$ , får vi som et biresultat at vide hvad  $b(j, k)$  er. Det ser pænere ud når man skriver op hvad  $b(2x+2, 2y+2)$  er, så her er de to udtryk for  $b(2x+2, 2y+2)$  der er lig med hinanden<sup>d</sup>:

$$\int_0^1 t^x (1-t)^y dt = \frac{x! y!}{(x+y)!}, \quad x, y > -1.$$

### 4.3 Summer af normalfordelte variable

Efter at have studeret summer af kvadrater på  $\mathcal{N}(0, 1)$ -variable vil vi nu se på almindelige summer af normalfordelte stokastiske variable. Der gælder

<sup>d</sup>Af hensyn til almendannelsen kan vi nævne, at den såkaldte *betafunktion* (eller *B-funktion*) defineres som funktionen

$$\begin{aligned}B(x, y) &= \int_0^1 t^{x-1} (1-t)^{y-1} dt \\ &= \frac{(x-1)! (y-1)!}{(x+y-2)!}, \quad x, y > 0.\end{aligned}$$

Hvis  $X_1, X_2, \dots, X_n$  er uafhængige normalfordelte stokastiske variable (gerne med hver sin middelværdi og varians), så er summen

$$S = X_1 + X_2 + \dots + X_n$$

normalfordelt.

Der gælder (som altid) at middelværdien af  $S$  er lig summen af de enkelte variables middelværdier, og (som altid for uafhængige variable) at variansen af  $S$  er lig summen af de enkelte variables varians.

Ved at kombinere dette resultat med resultatet om affine transformationer af normalfordelingen (side 69) fås

Hvis  $X_1, X_2, \dots, X_n$  er uafhængige *identisk*  $\mathcal{N}(\mu, \sigma^2)$ -fordelte stokastiske variable, så er gennemsnittet  $\bar{X}$  normalfordelt med middelværdi  $\mu$  og varians  $\sigma^2/n$ .

– Når man skal bevise resultater om fordelingen af  $S$  er det nok at gøre det for  $n = 2$ . Det er også nok at se på tilfældet  $\mu_1 = \mu_2 = 0$ , fordi man kan altid omskrive  $X_1 + X_2$  til  $\mu_1 + \mu_2 + (X_1 - \mu_1) + (X_2 - \mu_2)$ , hvor jo  $(X_i - \mu_i)$  er  $\mathcal{N}(0, \sigma_i^2)$ -fordelt. Lad derfor  $X_1$  og  $X_2$  være stokastisk uafhængige normalfordelte variable med middelværdi 0 og varianser  $\sigma_1^2$  og  $\sigma_2^2$ . Tætheden for  $X_1 + X_2$  er da (jf. (2.12) side 47)

$$g(y) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \frac{1}{\sqrt{2\pi\sigma_2^2}} \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2}\left[\frac{z^2}{\sigma_1^2} + \frac{(y-z)^2}{\sigma_2^2}\right]\right) dz,$$

der ved udvisning af tilstrækkelig meget tålmodighed og stædighed let kan omskrives til

$$\frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}} \exp\left(-\frac{1}{2}\frac{y^2}{\sigma_1^2 + \sigma_2^2}\right) \times \\ \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}}} \exp\left(-\frac{1}{2}\frac{\left(z - \frac{\sigma_1^2 y}{\sigma_1^2 + \sigma_2^2}\right)^2}{\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}}\right) dz,$$

hvor integralet simpelthen er integralet af en normalfordelingstæthed og derfor er lig 1. Tætheden for  $X_1 + X_2$  er dermed

$$g(y) = \frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}} \exp\left(-\frac{1}{2} \frac{y^2}{\sigma_1^2 + \sigma_2^2}\right),$$

hvilket som ønsket er tætheden for normalfordelingen med middelværdi 0 og varians  $\sigma_1^2 + \sigma_2^2$ .

## 4.4 Fordelingen af $\bar{X}$ og $s^2$

Man kan argumentere for, jf. statistiknoterne, at når  $x_1, x_2, \dots, x_n$  er observationer af indbyrdes uafhængige identisk  $\mathcal{N}(\mu, \sigma^2)$ -fordelte stokastiske variable  $X_1, X_2, \dots, X_n$ , så er tallene  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$  og  $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$  ræsonnable bud på ("estimer for") parametrene  $\mu$  og  $\sigma^2$ . Man beskæftiger sig derfor med spørgsmålet om, hvad fordelingen er af de to transformerede stokastiske variable

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

og

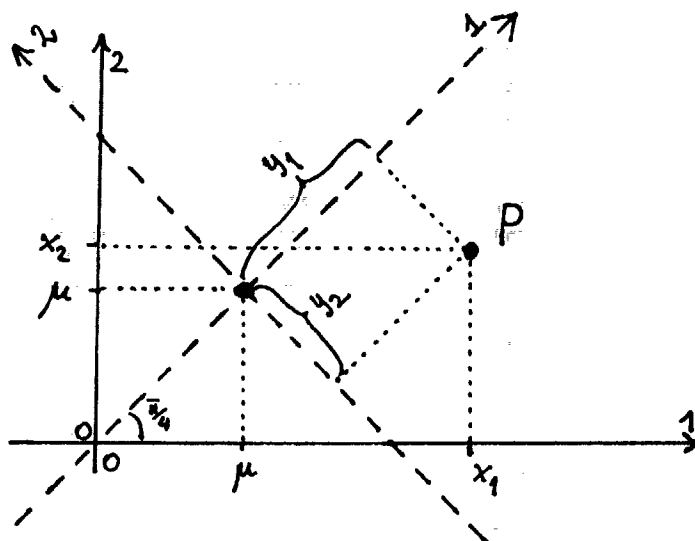
$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Man kan vise følgende vigtige resultat:

Hvis  $X_1, X_2, \dots, X_n$  er uafhængige  $\mathcal{N}(\mu, \sigma^2)$ -fordelte stokastiske variable, så gælder (for  $n > 1$ )

1.  $\bar{X}$  er  $\mathcal{N}(\mu, \frac{\sigma^2}{n})$ -fordelt.
2.  $s^2$  er  $\frac{\sigma^2}{n-1} \chi^2$ -fordelt med  $n - 1$  frihedsgrader.
3.  $\bar{X}$  og  $s^2$  er stokastisk uafhængige.

(Udsagnet i 2. er en anden formulering af udsagnet " $\frac{n-1}{\sigma^2} s^2$  er  $\chi^2$ -fordelt med  $n - 1$  frihedsgrader".)



**Figur 4.4:** Det nye og det gamle koordinatsystem. Det gamle koordinatsystems akser er fuldt optrukne, det nyes er stiplede.

Vi har allerede set at 1. gælder. Vi vil bevise 2. og 3. i tilfældet  $n = 2$ . Det handler her (som i det generelle tilfælde) om at vælge en passende transformation af  $X$ -erne. Vi betragter en transformation  $(y_1, y_2) = t(x_1, x_2)$  svarende til at vi skifter koordinatsystem. I det gamle koordinatsystem har punkter koordinater  $(x_1, x_2)$ . Det nye koordinatsystem, hvor koordinaterne hedder  $(y_1, y_2)$ , har sit begyndelsespunkt i  $(x_1, x_2) = (\mu, \mu)$  og dets akser er drejet vinklen  $\pi/4$  i forhold til det gamle koordinatsystem, se Figur 4.4. Hvis et punkt  $P$  har koordinater  $(x_1, x_2)$  i det gamle koordinatsystem og koordinater  $(y_1, y_2)$  i det nye, så gælder at  $(x_1 - \mu)^2 + (x_2 - \mu)^2 = y_1^2 + y_2^2$  er den kvadratiske afstand mellem  $P$  og det nye koordinatsystems begyndelsespunkt. Desuden er  $y_1 = \sqrt{2}(\bar{x} - \mu)$ , og dermed  $y_2^2 = (y_1^2 + y_2^2) - y_1^2 = (x_1 - \mu)^2 + (x_2 - \mu)^2 - 2(\bar{x} - \mu)^2 = (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 = s^2$ .

$\bar{x}$  og  $s^2$  er altså "pæne" funktioner af  $y_1$  og  $y_2$ , nemlig

$$\bar{x} = \mu + \frac{y_1}{\sqrt{2}}$$

$$s^2 = y_2^2.$$

Vi vil finde fordelingen af de af  $X_1$  og  $X_2$  transformerende stokastiske variable  $Y_1$  og  $Y_2$ . Da transformationen blot består i at flytte rundt med



koordinatsystemet, er "volumenforholdet" (som her hvor  $n = 2$  er et arealforhold) lig 1. Tætheden for  $(Y_1, Y_2)$  er derfor  $X$ -ernes simultane tæthed

$$\frac{1}{\sqrt{2\pi}\sigma^2} \exp\left(-\frac{1}{2} \frac{(x_1 - \mu)^2 + (x_2 - \mu)^2}{\sigma^2}\right)$$

hvor  $x_1$  og  $x_2$  nu betyder  $x_1$  og  $x_2$  som funktion af  $y_1$  og  $y_2$ . Men som nævnt er  $(x_1 - \mu)^2 + (x_2 - \mu)^2 = y_1^2 + y_2^2$ , så tætheden for  $(Y_1, Y_2)$  er

$$\begin{aligned} & \frac{1}{2\pi\sigma^2} \exp\left(-\frac{1}{2} \frac{y_1^2 + y_2^2}{\sigma^2}\right) \\ &= \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left(-\frac{1}{2} \frac{y_1^2}{\sigma^2}\right) \times \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left(-\frac{1}{2} \frac{y_2^2}{\sigma^2}\right). \end{aligned}$$

Det ses heraf at  $Y_1$  og  $Y_2$  er stokastisk uafhængige  $\mathcal{N}(0, \sigma^2)$ -variable. Da  $Y_1$  og  $Y_2$  er uafhængige, er  $\bar{X} = \mu + Y_1/\sqrt{2}$  og  $s^2 = Y_2^2$  også uafhængige, hvilket var påstand 3. Endelig følger påstand 2. af at  $s^2/\sigma^2 = (Y_2/\sigma)^2$  hvor  $Y_2/\sigma$  er  $\mathcal{N}(0, 1)$ -fordelt, og at kvadratet på en  $\mathcal{N}(0, 1)$ -variabel er  $\chi^2$ -fordelt med én frihedsgrad.

– Med et lidt mere avanceret geometrisk begrebsapparat til sin rådighed er det overordentlig simpelt at bevise sætningen for vilkårligt  $n$ , men det skal vi ikke komme ind på her.

## 4.5 $t$ -fordelingen, $F$ -fordelingen

Afsluttende skal kort omtales to teststørrelse-fordelinger der ofte benyttes, nemlig  $t$ -fordelingen og  $F$ -fordelingen.

$t$ -fordelingen er fordelingen af en brøk hvor tælleren er en  $\mathcal{N}(0, \tau^2)$ -

variabel og nævneren et af tælleren stokastisk uafhængigt skøn over  $\tau^2$ :

Hvis

- $Y$  er  $\mathcal{N}(0, \tau^2)$ -fordelt, og
- $Z$  er  $\frac{\tau^2}{k} \chi^2$ -fordelt med  $k$  frihedsgrader, og
- $Y$  og  $Z$  er stokastisk uafhængige,

så er

$$t = \frac{Y}{\sqrt{Z}}$$

$t$ -fordelt med  $k$  frihedsgrader.

#### Eksempel 4.1. $t$ -teststørrelsen

Hvis  $X_1, X_2, \dots, X_n$  er stokastisk uafhængige  $\mathcal{N}(\mu, \sigma^2)$ -fordelte, så er

$$t = \frac{\bar{X} - \mu}{\sqrt{s^2/n}}$$

$t$ -fordelt med  $n - 1$  frihedsgrader. Thi fra side 81 ved vi, at  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  er normalfordelt med middelværdi  $\mu$  og varians  $\tau^2 = \sigma^2/n$ , og at  $s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$  er  $\frac{\sigma^2}{n-1} \chi^2$ -fordelt og stokastisk uafhængig af  $\bar{X}$ .  $\square$

$F$ -fordelingen er fordelingen af forholdet mellem to uafhængige  $\chi^2$ -

størrelser der er skøn over den samme variansparameter  $\sigma^2$ :

Hvis

- $Y$  er  $\frac{\sigma^2}{t}\chi^2$ -fordelt med  $t$  frihedsgrader,
- $Z$  er  $\frac{\sigma^2}{n}\chi^2$ -fordelt med  $n$  frihedsgrader, og
- $Y$  og  $Z$  er stokastisk uafhængige,

så er

$$f = \frac{Y}{Z}$$

*F-fordelt med frihedsgradsantal  $(t,n)$ .*

*F-fordelingen kaldes også  $v^2$ -fordelingen.*



# Liste over eksempler

1.1	Terningekast . . . . .	5
1.2	Terningekast, fortsat . . . . .	9
1.3	Terningekast, fortsat . . . . .	14
1.4	Marginal og simultan fordeling . . . . .	15
1.5	Opbygning af større modeller ud fra mindre . . . . .	18
1.6	Om fortolkningen af usædvanlige hændelser . . . . .	19
2.1	Den geometriske fordeling . . . . .	23
2.2	Fordelingsfunktionen for den geometriske fordeling . . . . .	29
2.3	Eksponentialfordelingen . . . . .	31
2.4	Eksempel på bestemmelse af fordelingen af summen $Y = X_1 + X_2$ af to heltallige stokastiske variable $X_1$ og $X_2$ . . . . .	37
2.5	Transformation af ligefordeling til eksponentialforde- ling. Simulation af eksponentialfordelte variable . . . . .	41
2.6	En sum af to eksponentialfordelte variable . . . . .	44
2.7	Transformation ved hjælp af polære koordinater . . . . .	48
2.8	Fordelingen af to eksponentialfordelte variable givet de- res sum . . . . .	51
3.1	En fordeling uden middelværdi . . . . .	55
4.1	$t$ -teststørrelsen . . . . .	84

# Stikordsregister

- affin transformation 69
- afhængighed, stokastisk 16
- antalsvariabel 21
- B*-funktion 79
- betafunktion 79
- betinget fordeling 7, 10, 12
- betinget sandsynlighed 12
- betingning 10
- Cauchy-fordeling 59
- diskret 21
- eksponentialfordeling 31, 43
- F*-fordeling 85
- fakultetsfunktion 75
- flerdimensional stokastisk variabel 14
- fordelingsfunktion 28
- fortolkning af sandsynlighed 2, 19
- forventet værdi 54
- fraktil 31
- frihedsgrader 75
- funktional 62
- funktionaldeterminant 44
- gammafordeling 46
- gammafunktion 77
- geometrisk fordeling 23, 29
- h*-interval 25, 35
- hypotetisk uendelig population 2
- hyppighed i det lange løb 2
- hændelse 4
- intensitet 31
- kontinuert fordeling 24
- kontinuert variabel 21
- korrelation 65
- kovarians 65
- kumuleret fordelingsfunktion 28
- kvadratisk skalaparameter 69
- kvalitativ 21
- kvantitativ 21
- ligefordeling 5, 42
- lineær funktional 62
- lotteri 54
- marginal fordeling 15
- marginal tæthed 36
- middelværdi 53
- n*-dimensional stokastisk variabel 14
- n*-tupel 35
- normalfordeling 67
- normeret funktional 62
- normeret normal fordeling 69

- polære koordinater 48
- positionsparameter 69
- positiv funktional 62
- sandsynlighedsfordeling 4, 5
- sandsynlighedsfunktion 4, 15
- sandsynlighedstæthedsfunktion 25, 35
- simulation af eksponentialfordelte tal 43
- simultan fordeling 15
- skalaækvivariant 64
- standardafvigelse 63
- stokastisk afhængighed 16
- stokastisk uafhængighed 16
- stokastisk variabel 6, 14
- Store Tals Lov 1, 55
- summer af stokastiske variable 37
- t*-fordeling 84
- teststørrelsefordeling 83
- tilfældige tal 43
- transformation 7
- transformation af sandsynlighedsfordeling 7
- transformation af tætheder 39
- translationsinvariant 64
- translationsækvivariant 63
- tælletal 21
- uafhængige gentagelser 55
- uafhængighed, stokastisk 16
- udfald 3
- udfaldsrum 3
- $v^2$ -fordeling 85
- varians 63
- ventetid 31
- $\alpha$ -fraktil 33
- $\chi^2$ -fordeling 75, 76
- $\Gamma$ -funktion 77





- 1/78 "TANKER OM EN PRAKSIS" - et matematikprojekt. Projektrapport af: Anne Jensen, Lena Lindenskov, Marianne Kesselhahn og Nicolai Lomholt. Vejleder: Anders Madsen
- 2/78 "OPTIMERING" - Menneskets forøgede beherskelsermuligheder af natur og samfund. Projektrapport af: Tom J. Andersen, Tommy R. Andersen, Gert Krenøe og Peter H. Lassen. Vejleder: Bernhelm Boss.
- 3/78 "OPCAVESAMLING", breddekursus i fysik. Af: Lasse Rasmussen, Aage Bonde Kræmmer og Jens Højgaard Jensen.
- 4/78 "TRE ESSAYS" - om matematikundervisning, matematiklæreruddannelsen og videnskabsrindalismen. Af: Mogens Niss. Nr. 4 er p.t. udgået.
- 5/78 "BIBLIOGRAFISK VEJLEDNING til studiet af DEN MODERNE FYSIKS HISTORIE". Af: Helge Kragh. Nr. 5 er p.t. udgået.
- 6/78 "NOGLE ARTIKLER OG DEBATINDLÆG OM - læreruddannelse og undervisning i fysik, og - de naturvidenskabelige fags situation efter studenteroprøret". Af: Karin Beyer, Jens Højgaard Jensen og Bent C. Jørgensen.
- 7/78 "MATEMATIKKENS FORHOLD TIL SAMFUNDSØKONOMIEN". Af: B.V. Gnedenko. Nr. 7 er udgået.
- 8/78 "DYNAMIK OG DIAGRAMMER", Introduktion til energy-bond-graph formalismen. Af: Peder Voetmann Christiansen.
- 9/78 "OM PRAKSIS' INDFLYDELSE PÅ MATEMATIKKENS UDVIKLING". - Motiver til Kepler's: "Nova Stereometria Doliorum Vinariorum". Projektrapport af: Lasse Rasmussen. Vejleder: Anders Madsen.
- 
- 10/79 "TERMODYNAMIK I GYMNASIET". Projektrapport af: Jan Christensen og Jeanne Mortensen. Vejledere: Karin Beyer og Peder Voetmann Christiansen.
- 11/79 "STATISTISKE MATERIALER". Af: Jørgen Larsen.
- 12/79 "LINEÆRE DIFFERENTIALLIGNINGER OG DIFFERENTIALLIGNINGSSYSTEMER". Af: Mogens Brun Heefelt. Nr. 12 er udgået.
- 13/79 "CAVENDISH'S FORSØG I GYMNASIET". Projektrapport af: Gert Kreinøe. Vejleder: Albert Chr. Paulsen.
- 14/79 "BOOKS ABOUT MATHEMATICS: History, Philosophy, Education, Models, System Theory, and Works of". Af: Else Høyrup. Nr. 14 er p.t. udgået.
- 15/79 "STRUKTUREL STABILITET OG KATASTROFER i systemer i og udenfor termodynamisk ligevægt". Specialeopgave af: Leif S. Striegler. Vejleder: Peder Voetmann Christiansen.
- 16/79 "STATISTIK I KRÆFTFORSKNINGEN". Projektrapport af: Michael Olsen og Jørn Jensen. Vejleder: Jørgen Larsen.
- 17/79 "AT SPØRGE OG AT SVARE i fysikundervisningen". Af: Albert Christian Paulsen.
- 18/79 "MATHEMATICS AND THE REAL WORLD", Proceedings af an International Workshop, Roskilde University Centre, Denmark, 1978. Preprint. Af: Bernhelm Booss og Mogens Niss (eds.)
- 19/79 "GEOMETRI, SKOLE OG VIRKELIGHED". Projektrapport af: Tom J. Andersen, Tommy R. Andersen og Per H.H. Larsen. Vejleder: Mogens Niss.
- 20/79 "STATISTISKE MODELLER TIL BESTEMMELSE AF SIKRE DOSER FOR CARCINOGENE STOFFER". Projektrapport af: Michael Olsen og Jørn Jensen. Vejleder: Jørgen Larsen.
- 21/79 "KONTROL I GYMNASIET-FORMÅL OG KONSEKVENSER". Projektrapport af: Crilles Bacher, Per S. Jensen, Preben Jensen og Torben Nysteen.
- 22/79 "SEMIOTIK OG SYSTEMEGENSKABER (1)". 1-port lineær response og støj i fysikken. Af: Peder Voetmann Christiansen.
- 23/79 "ON THE HISTORY OF EARLY WAVE MECHANICS - with special emphasis on the role of reality". Af: Helge Kragh.
- 
- 24/80 "MATEMATIKOPFATTELSE HOS 2.C'ERE". a+b 1. En analyse. 2. Interviewmateriale. Projektrapport af: Jan Christensen og Knud Lindhardt Rasmussen. Vejleder: Mogens Niss.
- 25/80 "EKSAMENSOPGAVER", Dybdemodul/fysik 1974-79.
- 26/80 "OM MATEMATISKE MODELLER". En projektrapport og to artikler. Af: Jens Højgaard Jensen m.fl.
- 27/80 "METHODOLOGY AND PHILOSOPHY OF SCIENCE IN PAUL DIRAC'S PHYSICS". Af: Helge Kragh.
- 28/80 "DILENTRISK RELAXATION - et forslag til en ny model bygget på væskernes viscoelastiske egenskaber". Projektrapport af: Gert Kreinøe. Vejleder: Niels Boye Olsen.
- 29/80 "ODIN - undervisningsmateriale til et kursus i differentialligningsmodeller". Projektrapport af: Tommy R. Andersen, Per H.H. Larsen og Peter H. Lassen. Vejleder: Mogens Brun Heefelt.
- 30/80 "FUSIONSENERGIEN - - - ATOMSAMFUNDETS ENDESTATION". Af: Oluf Danielsen. Nr. 30 er udgået.
- 31/80 "VIDENSKABSTEORETISKE PROBLEMER VED UNDERVISNINGSSYSTEMER BASERET PÅ Mængdelære". Projektrapport af: Troels Lange og Jørgen Karrebæk. Vejleder: Stig Andur Pedersen. Nr. 31 er p.t. udgået.
- 32/80 "POLYMERE STOFFERS VISCOELASTISKE EGENSKABER - BELYST VED HJÆLP AF MEKANISKE IMPEDANSMÅL - GER MOSSBAUEREFFEKT MÅLINGER". Projektrapport af: Crilles Bacher og Preben Jensen. Vejledere: Niels Boye Olsen og Peder Voetmann Christiansen.
- 33/80 "KONSTITUERING AF FAG INDEN FOR TEKNISK - NATURVIDENSKABELIGE UDDANNELSER. I-II". Af: Arne Jakobsen.
- 34/80 "ENVIRONMENTAL IMPACT OF WIND ENERGY UTILIZATION". ENERGY SERIES NO. 1. Af: Bent Sørensen. Nr. 34 er udgået.

- 35/80 "HISTORISKE STUDIER I DEN NYERE ATOMFYSIKS UDVIKLING".  
Af: Helge Kragh.
- 36/80 "HVAD ER MENINGEN MED MATEMATIKUNDERVISNINGEN?".  
Fire artikler.  
Af: Mogens Niss.
- 37/80 "RENEWABLE ENERGY AND ENERGY STORAGE".  
ENERGY SERIES NO. 2.  
Af: Bent Sørensen.
- 
- 38/81 "TIL EN HISTORIETORI OM NATURERKENDELSE, TEKNOLOGI OG SAMFUND".  
Projektrapport af: Erik Gade, Hans Hedal, Henrik Lau og Finn Physant.  
Vejledere: Stig Andur Pedersen, Helge Kragh og Ib Thiersen.  
Nr. 38 er p.t. udgået.
- 39/81 "TIL KRITIKKEN AF VÆKSTØKONOMIEN".  
Af: Jens Højgaard Jensen.
- 40/81 "TELEKOMMUNIKATION I DANMARK - oplæg til en teknologivurdering".  
Projektrapport af: Arne Jørgensen, Bruno Petersen og Jan Vedde.  
Vejleder: Per Nørgaard.
- 41/81 "PLANNING AND POLICY CONSIDERATIONS RELATED TO THE INTRODUCTION OF RENEWABLE ENERGY SOURCES INTO ENERGY SUPPLY SYSTEMS".  
ENERGY SERIES NO. 3.  
Af: Bent Sørensen.
- 42/81 "VIDENSKAB TEORI SAMFUND - En introduktion til materialistiske videnskabsopfattelser".  
Af: Helge Kragh og Stig Andur Pedersen.
- 43/81 1. "COMPARATIVE RISK ASSESSMENT OF TOTAL ENERGY SYSTEMS".  
2. "ADVANTAGES AND DISADVANTAGES OF DECENTRALIZATION".  
ENERGY SERIES NO. 4.  
Af: Bent Sørensen.
- 44/81 "HISTORISKE UNDERSØGELSER AF DE EKSPERIMENTELLE FORUDSÆTNINGER FOR RUTHERFORDS ATOMMODEL".  
Projektrapport af: Niels Thor Nielsen.  
Vejleder: Bent C. Jørgensen.
- 
- 45/82 Er aldrig udkommet.
- 46/82 "EKSEMPLARISK UNDERVISNING OG FYSISK ERKENDELSE-1+11 ILLUSTRERET VED TO EKSEMPLER".  
Projektrapport af: Torben O. Olsen, Lasse Rasmussen og Niels Dreyer Sørensen.  
Vejleder: Bent C. Jørgensen.
- 47/82 "BARSEBÄCK OG DET VÆRST OFFICIELT-TÆNKELIGE UHELD".  
ENERGY SERIES NO. 5.  
Af: Bent Sørensen.
- 48/82 "EN UNDERSØGELSE AF MATEMATIKUNDERVISNINGEN PÅ ADGANGSKURSUS TIL KØBENHAVNS TEKNIKUM".  
Projektrapport af: Lis Eilertzen, Jørgen Karrebæk, Troels Lange, Preben Nørregaard, Lissi Pedersen, Laust Rishøj, Lill Røn og Isac Showiki.  
Vejleder: Mogens Niss.
- 49/82 "ANALYSE AF MULTISPEKTRALE SATELLITBILLEDER".  
Projektrapport af: Preben Nørregaard.  
Vejledere: Jørgen Larsen og Rasmus Ole Rasmussen.
- 50/82 "HERSLEV - MULIGHEDER FOR VEDVARENDE ENERGI I EN LANDSBY".  
ENERGY SERIES NO. 6.  
Rapport af: Bent Christensen, Bent Hove Jensen, Dennis B. Møller, Bjarne Laursen, Bjarne Lillethorup og Jacob Mørch Pedersen.  
Vejleder: Bent Sørensen.
- 51/82 "HVAD KAN DER GØRES FOR AT AFHJÆLPE PIGERS BLOKERING OVERFOR MATEMATIK?".  
Projektrapport af: Lis Eilertzen, Lissi Pedersen, Lill Røn og Susanne Stender.
- 52/82 "DESUSPENSION OF SPLITTING ELLIPTIC SYMBOLS".  
Af: Bernhelm Booss og Krzysztof Wojciechowski.
- 53/82 "THE CONSTITUTION OF SUBJECTS IN ENGINEERING EDUCATION".  
Af: Arne Jacobsen og Stig Andur Pedersen.
- 54/82 "FUTURES RESEARCH" - A Philosophical Analysis of Its Subject-Matter and Methods.  
Af: Stig Andur Pedersen og Johannes Witt-Hansen.
- 55/82 "MATEMATISKE MODELLER" - Litteratur på Roskilde Universitetsbibliotek.  
En biografi.  
Af: Else Højrup.  
  
Vedr. tekst nr. 55/82 se også tekst nr. 62/83.
- 56/82 "EN - TO - MANGE" -  
En undersøgelse af matematisk økologi.  
Projektrapport af: Troels Lange.  
Vejleder: Anders Madsen.
- 
- 57/83 "ASPECT EKSPERIMENTET" -  
Skjulte variable i kvantemekanikken?  
Projektrapport af: Tom Juul Andersen.  
Vejleder: Peder Voetmann Christiansen.  
Nr. 57 er udgået.
- 58/83 "MATEMATISKE VANDRINGER" - Modelbetragtninger over spredning af dyr mellem småbiotoper i agerlandet.  
Projektrapport af: Per Hammershøj Jensen og Lene Vagn Rasmussen.  
Vejleder: Jørgen Larsen.
- 59/83 "THE METHODOLOGY OF ENERGY PLANNING".  
ENERGY SERIES NO. 7.  
Af: Bent Sørensen.
- 60/83 "MATEMATISK MODEKSPERTISE" - et eksempel.  
Projektrapport af: Erik O. Gade, Jørgen Karrebæk og Preben Nørregaard.  
Vejleder: Anders Madsen.
- 61/83 "FYSIKS IDEOLOGISKE FUNKTION, SOM ET EKSEMPEL PÅ EN NATURVIDENSKAB - HISTORISK SET".  
Projektrapport af: Annette Post Nielsen.  
Vejledere: Jens Højrup, Jens Højgaard Jensen og Jørgen Vogelius.
- 62/83 "MATEMATISKE MODELLER" - Litteratur på Roskilde Universitetsbibliotek.  
En biografi 2. rev. udgave.  
Af: Else Højrup.
- 63/83 "CREATING ENERGY FUTURES: A SHORT GUIDE TO ENERGY PLANNING".  
ENERGY SERIES No. 8.  
Af: David Crossley og Bent Sørensen.
- 64/83 "VON MATEMATIK UND KRIEG".  
Af: Bernhelm Booss og Jens Højrup.
- 65/83 "ANVENDT MATEMATIK - TEORI ELLER PRAKSIS".  
Projektrapport af: Per Hødegård Andersen, Kirsten Habekost, Carsten Holst-Jensen, Annelise von Moos, Else Marie Pedersen og Erling Møller Pedersen.  
Vejledere: Bernhelm Booss og Klaus Grünbaum.
- 66/83 "MATEMATISKE MODELLER FOR PERIODISK SELEKTION I ESCHERICHIA COLI".  
Projektrapport af: Hanne Lisbet Andersen, Ole Richard Jensen og Klavs Frisdahl.  
Vejledere: Jørgen Larsen og Anders Hede Madsen.
- 67/83 "ELEPSOIDE METODEN - EN NY METODE TIL LINEÆR PROGRAMMERING?".  
Projektrapport af: Lone Billmann og Lars Boye.  
Vejleder: Mogens Brun Heefelt.
- 68/83 "STOKASTISKE MODELLER I POPULATIONSGENETIK" - til kritikken af teoriladede modeller.  
Projektrapport af: Lise Odgård Gade, Susanne Hansen, Michael Hviid og Frank Mølgård Olsen.  
Vejleder: Jørgen Larsen.

- 69/83 "ELEVFORUDSÆTNINGER I FYSIK"  
- en test i l.g med kommentarer.  
Af: Albert C. Paulsen.
- 70/83 "INDLÆRINGS - OG FORMIDLINGSPROBLEMER I MATEMATIK PÅ VOKSENUNDERVISNINGSNIVEAU".  
Projektrapport af: Hanne Lisbet Andersen, Torben J. Andreassen, Svend Åge Houmann, Helle Glerup Jensen, Keld Fl. Nielsen, Lene Vagn Rasmussen.  
Vejleder: Klaus Grünbaum og Anders Hede Madsen.
- 71/83 "PIGER OG FYSIK"  
- et problem og en udfordring for skolen?  
Af: Karin Beyer, Sussanne Blegaa, Birthe Olsen, Jette Reich og Mette Vedelsby.
- 72/83 "VERDEN IFVLGE PEIRCE" - to metafysiske essays, om og af C.S Peirce.  
Af: Peder Voetmann Christiansen.
- 73/83 "EN ENERGIANALYSE AF LANDBRUG"  
- økologisk contra traditionelt.  
ENERGY SERIES NO. 9  
Specialeopgave i fysik af: Bent Hove Jensen.  
Vejleder: Bent Sørensen.
- 74/84 "MINIATURISERING AF MIKROELEKTRONIK" - om videnskabeliggjort teknologi og nytten af at lære fysik.  
Projektrapport af: Bodil Harder og Linda Szkotak Jensen.  
Vejledere: Jens Højgaard Jensen og Bent C. Jørgensen.
- 75/84 "MATEMATIKUNDERVISNINGEN I FREMTIDENS GYMNASIUM"  
- Case: Lineær programmering.  
Projektrapport af: Morten Blomhøj, Klavs Frisdahl og Frank Mølgaard Olsen.  
Vejledere: Mogens Brun Heefelt og Jens Bjørneboe.
- 76/84 "KERNEKRAFT I DANMARK?" - Et høringssvar indkaldt af miljøministeriet, med kritik af miljøstyrelsens rapporter af 15. marts 1984.  
ENERGY SERIES No. 10  
Af: Niels Boye Olsen og Bent Sørensen.
- 77/84 "POLITISKE INDEKS - FUP ELLER FAKTA?"  
Opinionsundersøgelser belyst ved statistiske modeller.  
Projektrapport af: Svend Åge Houmann, Keld Nielsen og Susanne Stender.  
Vejledere: Jørgen Larsen og Jens Bjørneboe.
- 78/84 "JÆVNSTRØMSLEDNINGSEVNE OG GITTERSTRUKTUR I AMORFT GERMANIUM".  
Specialrapport af: Hans Heddal, Frank C. Ludvigsen og Finn C. Physant.  
Vejleder: Niels Boye Olsen.
- 79/84 "MATEMATIK OG ALMENDANNELSE".  
Projektrapport af: Henrik Ooster, Mikael Wennerberg Johansen, Povl Kattler, Birgitte Lydholm og Morten Overgaard Nielsen.  
Vejleder: Bernhelm Booss.
- 80/84 "KURSUSMATERIALE TIL MATEMATIK B".  
Af: Mogens Brun Heefelt.
- 81/84 "FREKVENSafhængig ledningsevne i amorft germanium".  
Specialrapport af: Jørgen Wind Petersen og Jan Christensen.  
Vejleder: Niels Boye Olsen.
- 82/84 "MATEMATIK - OG FYSIKUNDERVISNINGEN I DET AUTOMATISEREDE SAMFUND".  
Rapport fra et seminar afholdt i Hvidovre 25-27 april 1983.  
Red.: Jens Højgaard Jensen, Bent C. Jørgensen og Mogens Niss.
- 83/84 "ON THE QUANTIFICATION OF SECURITY":  
PEACE RESEARCH SERIES NO. 1  
Af: Bent Sørensen  
nr. 83 er p.t. udgået
- 84/84 "NOGLE ARTIKLER OM MATEMATIK, FYSIK OG ALMENDANNELSE".  
Af: Jens Højgaard Jensen, Mogens Niss m. fl.
- 85/84 "CENTRIFUGALREGULATORER OG MATEMATIK".  
Specialrapport af: Per Hedegård Andersen, Carsten Holst-Jensen, Else Marie Pedersen og Erling Møller Pedersen.  
Vejleder: Stig Andur Pedersen.
- 86/84 "SECURITY IMPLICATIONS OF ALTERNATIVE DEFENSE OPTIONS FOR WESTERN EUROPE".  
PEACE RESEARCH SERIES NO. 2  
Af: Bent Sørensen.
- 87/84 "A SIMPLE MODEL OF AC HOPPING CONDUCTIVITY IN DISORDERED SOLIDS".  
Af: Jeppe C. Dyre.
- 88/84 "RISE, FALL AND RESURRECTION OF INFINITESIMALS".  
Af: Detlef Laugwitz.
- 89/84 "FUERNVARMEOPTIMERING".  
Af: Bjarne Lillethorup og Jacob Mørch Pedersen.
- 90/84 "ENERGI I l.G - EN TEORI FOR TILRETTELÆGGELSE".  
Af: Albert Chr. Paulsen.
- 91/85 "KVANTETEORI FOR GYMNASIET".  
1. Lærervejledning  
Projektrapport af: Biger Lundgren, Henning Sten Hansen og John Johansson.  
Vejleder: Torsten Meyer.
- 92/85 "KVANTETEORI FOR GYMNASIET".  
2. Materiale  
Projektrapport af: Biger Lundgren, Henning Sten Hansen og John Johansson.  
Vejleder: Torsten Meyer.
- 93/85 "THE SEMIOTICS OF QUANTUM - NON - LOCALITY".  
Af: Peder Voetmann Christiansen.
- 94/85 "TREENIGHEDEN BOURBAKI - generalen, matematikeren og ånden".  
Projektrapport af: Morten Blomhøj, Klavs Frisdahl og Frank M. Olsen.  
Vejleder: Mogens Niss.
- 95/85 "AN ALTERNATIV DEFENSE PLAN FOR WESTERN EUROPE".  
PEACE RESEARCH SERIES NO. 3  
Af: Bent Sørensen
- 96/85 "ASPEKTER VED KRAFTVARMEFORSYNING".  
Af: Bjarne Lillethorup.  
Vejleder: Bent Sørensen.
- 97/85 "ON THE PHYSICS OF A.C. HOPPING CONDUCTIVITY".  
Af: Jeppe C. Dyre.
- 98/85 "VALGMULIGHEDER I INFORMATIONSSALDEREN".  
Af: Bent Sørensen.
- 99/85 "Der er langt fra Q til R".  
Projektrapport af: Niels Jørgensen og Mikael Klintorp.  
Vejleder: Stig Andur Pedersen.
- 100/85 "TALSYSTEMETS OPBYGNING".  
Af: Mogens Niss.
- 101/85 "EXTENDED MOMENTUM THEORY FOR WINDMILLS IN PERTURBATIVE FORM".  
Af: Ganesh Sengupta.
- 102/85 OPSTILLING OG ANALYSE AF MATEMATISKE MODELLER, BELYST VED MODELLER OVER KØRS FODEROPFØDELSE OG - OMSÆTNING".  
Projektrapport af: Lis Eilertzen, Kirsten Habekost, Lill Røn og Susanne Stender.  
Vejleder: Klaus Grünbaum.

- 103/85 "ØDSLE KOLDKRIGERE OG VIDENSKABENS LYSE IDEER".  
Projektrapport af: Niels Ole Dam og Kurt Jensen.  
Vejleder: Bent Sørensen.
- 104/85 "ANALOGREGNEMASKINEN OG LORENZLIGNINGER".  
Af: Jens Jøger.
- 105/85 "THE FREQUENCY DEPENDENCE OF THE SPECIFIC HEAT OF THE GLASS TRANSITION".  
Af: Tage Christensen.
- "A SIMPLE MODEL OF AC HOPPING CONDUCTIVITY".  
Af: Jeppe C. Dyre.  
Contributions to the Third International Conference on the Structure of Non - Crystalline Materials held in Grenoble July 1985.
- 106/85 "QUANTUM THEORY OF EXTENDED PARTICLES".  
Af: Bent Sørensen.
- 107/85 "EN MYG GØR INGEN EPIDEMI".  
- flodblindhed som eksempel på matematisk modellering af et epidemiologisk problem.  
Projektrapport af: Per Hedegård Andersen, Lars Boye, Carsten Holst Jensen, Else Marie Pedersen og Erling Møller Pedersen.  
Vejleder: Jesper Larsen.
- 108/85 "APPLICATIONS AND MODELLING IN THE MATHEMATICS CURRICULUM" - state and trends -  
Af: Mogens Niss.
- 109/85 "COX I STUDIETIDEN" - Cox's regressionsmodel anvendt på studenteroplysninger fra RUC.  
Projektrapport af: Mikael Wennerberg Johansen, Poul Kattler og Torben J. Andreassen.  
Vejleder: Jørgen Larsen.
- 110/85 "PLANNING FOR SECURITY".  
Af: Bent Sørensen
- 111/85 "JORDEN RUNDT PÅ FLADE KORT".  
Projektrapport af: Birgit Andresen, Beatriz Quinones og Jimmy Staal.  
Vejleder: Mogens Niss.
- 112/85 "VIDENSKABELIGGØRELSE AF DANSK TEKNOLOGISK INNOVATION FREM TIL 1950 - BELYST VED EKSEMPLER".  
Projektrapport af: Erik Odgaard Gade, Hans Hødal, Frank C. Ludvigsen, Annette Post Nielsen og Finn Physant.  
Vejleder: Claus Bryld og Bent C. Jørgensen.
- 113/85 "DESUSPENSION OF SPLITTING ELLIPTIC SYMBOLS 11".  
Af: Bernhelm Booss og Krzysztof Wojciechowski.
- 114/85 "ANVENDELSE AF GRAFISKE METODER TIL ANALYSE AF KONTIGENTABELER".  
Projektrapport af: Lone Biilmann, Ole R. Jensen og Arne-Lise von Moos.  
Vejleder: Jørgen Larsen.
- 115/85 "MATEMATIKKENS UDVIKLING OP TIL RENESSANCEN".  
Af: Mogens Niss.
- 116/85 "A PHENOMENOLOGICAL MODEL FOR THE MEYER-NELDEL RULE".  
Af: Jeppe C. Dyre.
- 117/85 "KRAFT & FJERNVARMEOPTIMERING".  
Af: Jacob Mørch Pedersen.  
Vejleder: Bent Sørensen
- 118/85 "TILFÆLDIGHEDEN OG NØDVENDIGHEDEN IFØLGE PEIRCE OG FYSIKKEN".  
Af: Peder Voetmann Christiansen
- 119/86 "DET ER GANSKE VIST - - EUKLIDS FEMTE POSTULAT KUNNE NOK SKABE RØRE I ANDEDAMMEN".  
Af: Iben Maj Christiansen  
Vejleder: Mogens Niss.
- 120/86 "ET ANIAL STATISTISKE STANDARDMODELLER".  
Af: Jørgen Larsen
- 121/86 "SIMULATION I KONTINUERT TID".  
Af: Peder Voetmann Christiansen.
- 122/86 "ON THE MECHANISM OF GLASS IONIC CONDUCTIVITY".  
Af: Jeppe C. Dyre.
- 123/86 "GYMNASIEFYSIKKEN OG DEN STORE VERDEN".  
Fysiklærerforeningen, IMFUFA, RUC.
- 124/86 "OPGAVESAMLING I MATEMATIK".  
Samtlige opgaver stillet i tiden 1974-jan. 1986.
- 125/86 "UVBY, 8 - systemet - en effektiv fotometrisk spektral-klassifikation af B-, A- og F-stjerner".  
Projektrapport af: Birger Lundgren.
- 126/86 "OM UDVIKLINGEN AF DEN SPECIELLE RELATIVITETSTEORI".  
Projektrapport af: Lise Odgaard & Linda Szkotak Jensen  
Vejledere: Karin Beyer & Stig Andur Pedersen.
- 127/86 "GALOIS' BIDRAG TIL UDVIKLINGEN AF DEN ABSTRAKTE ALGEBRA".  
Projektrapport af: Pernille Sand, Heine Larsen & Lars Frandsen.  
Vejleder: Mogens Niss.
- 128/86 "SMÅKRYB" - om ikke-standard analyse.  
Projektrapport af: Niels Jørgensen & Mikael Klintorp.  
Vejleder: Jeppe Dyre.
- 129/86 "PHYSICS IN SOCIETY"  
Lecture Notes 1983 (1986)  
Af: Bent Sørensen
- 130/86 "Studies in Wind Power"  
Af: Bent Sørensen
- 131/86 "FYSIK OG SAMFUND" - Et integreret fysik/historie-projekt om naturanskuelsens historiske udvikling og dens samfundsmæssige betingethed.  
Projektrapport af: Jakob Heckscher, Søren Brønd, Andy Wierød.  
Vejledere: Jens Høyrup, Jørgen Vogelius, Jens Højgaard Jensen.
- 132/86 "FYSIK OG DANNEELSE"  
Projektrapport af: Søren Brønd, Andy Wierød.  
Vejledere: Karin Beyer, Jørgen Vogelius.
- 133/86 "CHERNOBYL ACCIDENT: ASSESSING THE DATA. ENERGY SERIES NO. 15.  
AF: Bent Sørensen.
- 
- 134/87 "THE D.C. AND THE A.C. ELECTRICAL TRANSPORT IN AsSeTe SYSTEM"  
Authors: M.B.El-Den, N.B.Olsen, Ib Høst Pedersen,  
Petr Visčor
- 135/87 "INTUITIONISTISK MATEMATIKS METODER OG ERKENDELSES-TEORETISKE FORUDSÆTNINGER"  
MATEMATIKSPECIALE: Claus Larsen  
Vejledere: Anton Jensen og Stig Andur Pedersen
- 136/87 "Mystisk og naturlig filosofi: En skitse af kristendommens første og andet møde med græsk filosofi"  
Projektrapport af Frank Colding Ludvigsen  
Vejledere: Historie: Ib Thiersen  
Fysik: Jens Højgaard Jensen
- 137/87 "HOPMODELLER FOR ELEKTRISK LEDNING I UORDNEDE FASTE STOFFER" - Resume af licentiatafhandling  
Af: Jeppe Dyre  
Vejledere: Niels Boye Olsen og  
Peder Voetmann Christiansen.

138/87 "JOSEPHSON EFFECT AND CIRCLE MAP."

Paper presented at The International Workshop on Teaching Nonlinear Phenomena at Universities and Schools, "Chaos in Education". Balaton, Hungary, 26 April-2 May 1987.

By: Peder Voetmann Christiansen

139/87 "Machbarkeit nichtbeherrschbarer Technik durch Fortschritte in der Erkennbarkeit der Natur"

Af: Bernhelm Booss-Bavnbek  
Martin Bohle-Carbonell

140/87 "ON THE TOPOLOGY OF SPACES OF HOLOMORPHIC MAPS"

By: Jens Gravesen

141/87 "RADIOMETERS UDVIKLING AF BLODGASAPPARATUR - ET TEKNOLOGIHISTORISK PROJEKT"

Projektrapport af Finn C. Physant  
Vejleder: Ib Thiersen

142/87 "The Calderón Projektor for Operators With Splitting Elliptic Symbols"

by: Bernhelm Booss-Bavnbek og  
Krzysztof P. Wojciechowski

143/87 "Kursusmateriale til Matematik på NAT-BAS"

af: Mogens Brun Heefelt

144/87 "Context and Non-Locality - A Peircan Approach"

Paper presented at the Symposium on the Foundations of Modern Physics The Copenhagen Interpretation 60 Years after the Como Lecture. Joensuu, Finland, 6 - 8 august 1987.

By: Peder Voetmann Christiansen

145/87 "AIMS AND SCOPE OF APPLICATIONS AND MODELLING IN MATHEMATICS CURRICULA"

Manuscript of a plenary lecture delivered at ICMTA 3, Kassel, FRG 8.-11.9.1987

By: Mogens Niss

146/87 "BESTEMMELSE AF BULKRESISTIVITETEN I SILICIUM"

- en ny frekvensbaseret målemetode.

Fysikspeciale af Jan Vedde

Vejledere: Niels Boye Olsen & Petr Višćor

147/87 "Rapport om BIS på NAT-BAS"

redigeret af: Mogens Brun Heefelt

148/87 "Naturvidenskabsundervisning med Samfundsperspektiv"

af: Peter Colding-Jørgensen DLH  
Albert Chr. Paulsen

149/87 "In-Situ Measurements of the density of amorphous germanium prepared in ultra high vacuum"

by: Petr Višćor

150/87 "Structure and the Existence of the first sharp diffraction peak in amorphous germanium prepared in UHV and measured in-situ"

by: Petr Višćor

151/87 "DYNAMISK PROGRAMMERING"

Matematikprojekt af:  
Birgit Andresen, Keld Nielsen og Jimmy Staal  
Vejleder: Mogens Niss

152/87 "PSEUDO-DIFFERENTIAL PROJECTIONS AND THE TOPOLOGY OF CERTAIN SPACES OF ELLIPTIC BOUNDARY VALUE PROBLEMS"

by: Bernhelm Booss-Bavnbek  
Krzysztof P. Wojciechowski

153/88 "HALVLEDERTEKNOLOGIENS UDVIKLING MELLEM MILITÆRE OG CIVILE KRÆFTER"

Et eksempel på humanistisk teknologihistorie  
Historiespeciale

Af: Hans Hedal

Vejleder: Ib Thiersen

154/88 "MASTER EQUATION APPROACH TO VISCOUS LIQUIDS AND THE GLASS TRANSITION"

By: Jeppe Dyre

155/88 "A NOTE ON THE ACTION OF THE POISSON SOLUTION OPERATOR TO THE DIRICHLET PROBLEM FOR A FORMALLY SELFADJOINT DIFFERENTIAL OPERATOR"

by: Michael Pedersen

156/88 "THE RANDOM FREE ENERGY BARRIER MODEL FOR AC CONDUCTION IN DISORDERED SOLIDS"

by: Jeppe C. Dyre

157/88 "STABILIZATION OF PARTIAL DIFFERENTIAL EQUATIONS BY FINITE DIMENSIONAL BOUNDARY FEEDBACK CONTROL: A pseudo-differential approach."

by: Michael Pedersen

158/88 "UNIFIED FORMALISM FOR EXCESS CURRENT NOISE IN RANDOM WALK MODELS"

by: Jeppe Dyre

159/88 "STUDIES IN SOLAR ENERGY"

by: Bent Sørensen

160/88 "LOOP GROUPS AND INSTANTONS IN DIMENSION TWO"

by: Jens Gravesen

161/88 "PSEUDO-DIFFERENTIAL PERTURBATIONS AND STABILIZATION OF DISTRIBUTED PARAMETER SYSTEMS:

Dirichlet feedback control problems"

by: Michael Pedersen

162/88 "PIGER & FYSIK - OG MEGET MERE"

AF: Karin Beyer, Sussanne Blegaa, Birthe Olsen,  
Jette Reich, Mette Vedelsby

163/88 "EN MATEMATISK MODEL TIL BESTEMMELSE AF PERMEABILITETEN FOR BLOD-NETHINDE-BARRIEREN"

Af: Finn Langberg, Michael Jarden, Lars Frellesen  
Vejleder: Jesper Larsen